Evaluating the impact of road infrastructure on household income in Papua New Guinea

Koomen, E.; Andree, B.P.J.; Pradhan, M.P.; Wiegand, M.

2016

Evaluating the impact of road infrastructure on household income in Papua New Guinea

Spatial data compilation and analysis

Eric Koomen, Bo Andrée, Menno Pradhan and Martin Wiegand
December 20, 2016

VU VRIJE UNIVERSITEIT AMSTERDAM

Commissioned by ADB Asian Development Bank

Contract No. 117629-S85196 and 117642-S84962

TA-8332 REG: Developing Impact Evaluation Methodologies, Approaches, and Capacities in Selected Developing Member Countries (Subproject 1)

Front cover: road side view Eastern Highlands Province (by Eric Koomen October 10, 2015)

Contact info:
VU University Amsterdam
Faculty of Economics and Business Administration
Department of Spatial Economics/ Spatial Information Laboratory (SPINlab)
De Boelelaan 1105
1081 HV Amsterdam
Netherlands
Phone: +31 20 5986095
Email: e.koomen@vu.nl
Website: http://spinlab.vu.nl/

Summary

As part of the ADB sub-project 'Developing Impact Evaluation Methodologies, Approaches, and Capacities in Selected Developing Member Countries' two closely related research projects were carried out by VU to evaluate the impact of road infrastructure on household income in Papua New Guinea. One project focused on econometric analysis (contract no. 117629-S85196), while the other developed the spatial database needed for this analysis (contract no. 117642-S84962).

The current report documents the data collection process and provides a concise description of the data that were collected for the project and the analyses that were performed to enrich the available data sources to create meaningful variables for the statistical analysis. It serves as a background document with the scientific paper-that concisely describes the findings of the two joint research projects. The report is based on our initial progress reports and aims to provide more insight in the data sources that were available, the difficulties we encountered obtaining them and the analyses that were carried out to produce the basic variables used in the analysis.

One of the objectives of this report is to direct other scholars to the spatial data sources available for Papua New Guinea and hopefully limit the time they have to spend in collecting these data. The report also describes our efforts to explore new data sources (such as light-at-night or luminosity data and satellite imagery) and to construct new data sets (e.g. upgraded road stretches). While not all data sets were applied in the final econometric analysis, we document these efforts nonetheless to show what is feasible (and what not) with novel data sources. Another aim of this report is to provide some more explanation and background information on the spatially explicit data sets that were included in the eventual econometric analysis.

Contents

Introduction

Approved in 2009, the ADB financed Highlands Region Road Improvement and Investment Program (HRRIIP: 40173-023) focuses on upgrading the Highlands core road network (HCRN) of 2,500 kilometres (km) of major national and some provincial roads, which carry the bulk of the traffic in the highlands region of PNG. The investment program includes: (i) projects to improve about 1,400 km of the HCRN, to be funded through four tranches under this multi-tranche financing facility (MFF); and (ii) design and supervision of road improvement works, preparation and administration of long-term road maintenance contracts for the entire 2,500 km of the HCRN, and capacity development of road agencies. Under the program, two contracts have been awarded to develop total road length of 115km and construction of these roads is nearing completion. Contracts for an additional 120km of road section were approved in early 2014 with work to commence early 2015. HRRIIP continues more than a decade of ADB efforts to rehabilitate and maintain the HCRN with previous projects in the region including Highlands Road Rehabilitation and Maintenance (HRRM).

As part of the ADB sub-project 'Developing Impact Evaluation Methodologies, Approaches, and Capacities in Selected Developing Member Countries' two closely related research projects were carried out by VU to evaluate the impact of road infrastructure on household income in Papua New Guinea. One project focused on econometric analysis (contract no. 117629-S85196), while the other developed the spatial database needed for this analysis (contract no. 117642-S84962). This joint research effort is an update of an earlier study by Gibson and Rozelle (2003) that provided measures of how access to transport infrastructure, specifically rural feeder roads, impacted poverty in rural PNG. The current study makes use of a newer (2009/10) Household Income and Expenditure Survey data, and enriches the explanatory analysis with data obtained from new geo-referencing and space imaging (luminosity) techniques. The results of this study can be used to estimate the overall impact of ADB and Government road sector investments on poverty and welfare in the country and provide a benchmarking tool for the design of future investments into the region.

The analysis is performed at the household level and uses advanced econometric methods to show the relationship between poverty and access to roads, measured for example, as travelling time from a community to the nearest road. Poverty estimates are be combined with data on the quality of rural feeder roads during this period to the extent these are available. Potential sources for the latter type of information include data from Department of Transport data (including data on ADB road projects). As part of this project we also explored the possibility to obtain road quality data estimates from remote sensing images. In addition we obtained data sets describing 'luminosity' (light observed at night by satellites) in the highlands region to assess their potential to capture poverty and welfare in this data-poor country.

The econometric analysis that tested the impact of access to transport infrastructure on a household's poverty status between 1996 and 2010 is described in a separate paper that was prepared for submission to a refereed journal. The current report describes the spatial data collection process and discusses the data limitations that were encountered in the analysis. One of the objectives of this report is to direct other scholars to the spatial data sources available for Papua New Guinea and hopefully limit the time they have to spend in collecting these data. Anyone interested in receiving the data files mentioned in the report can contact the first author. The report, furthermore, describes our efforts to explore new data sources and to construct new data sets to show what is feasible (and what not) with these novel data sources. Another aim of this report is to provide some more explanation and background information on the spatially explicit data sets that were included in the econometric analysis.

Part I of this report describes the spatially explicit data that were collected to initiate the econometric analysis of the impact of road infrastructure on household income. The main

components of the collected spatial datasets are: administrative and population data, road network, altitude, luminosity, land cover and satellite imagery. As data availability for Papua New Guinea is limited, the database was initially filled with open access data sets from a variety of sources that cover larger parts of Oceania and Asia or even whole world. During the project additional data were added from local sources in Papua New Guinea and research colleagues in Australia and New Zealand.

Part II describes the information we were able to obtain on road improvement in the Highlands of Papua New Guinea following our field trip and subsequent desktop study. The final part of this report provides a fairy detailed, technical account of the spatial analyses that were eventually performed to produce the road-distance related explanatory variables for the econometric analysis.

As annex to this report we include the research paper that describes our attempts to recognise roads from available satellite imagery. The main purpose of these analyses was to assess the potential of remote-sensing based data to enrich the fairly incomplete records of road quality in the country. While we were able to develop an elaborate methodology to recognise main roads from Landsat images, we decided to not follow this data collection process as the recognition of more detailed roads was hampered by the limited availability of cloud-free, open data sources with sufficient spatial and temporal detail.

Part I Initial data compilation

1. Household survey data

The analysis on the impact of road construction in PNG will be based on the household surveys conducted in 1996 and 2009/2010. The census data of the 2000 and 2010 census may also provide useful information, in particular the poverty map that was constructed on the basis of the 1996 household survey and the 2000 census. We have access to the poverty map data and the household survey data. The 1996 household survey collects data from 1,144 households in 120 sampling clusters (47 urban, 73 rural). The 2009/2010 survey has a sample size of 4,081 households distributed over 321 clusters (196 urban, 125 rural). The sampling clusters are of particular importance in this analysis as most of the variation in road access will be between clusters, not within clusters.

Data on road access will be obtained from two sources. First, the household surveys contain information on proximity to roads and travel time to basic services. Second, we intend to use geographically tagged information on the road network, its location and quality, and how it changed in between the time of collection of the two household surveys. Our progress in compiling this data is described in more detail in subsequent sections of this report.

Table 1 provides national averages of some of the outcome variables that will be used in the analysis. Expenditure data and the resulting poverty estimates are hard to compare between the two surveys, which is partly due to changes in relative prices and the composition of the poverty line basket, but also because there were some fundamental differences in the data collection. In 1996 consumption data was collected via recall and using surveyed regional prices. In 2009/10 households had to keep consumption diaries, including self-reported prices, during a period of 14 days, with declining cooperation levels over time. For that reason we also consider a number of non-pecuniary indicators of well-being which are comparatively easily comparable between survey rounds. These include variables related to education, food shortages (only in 1996), being linked to the electricity grid (with at least four hours of electricity per day), access to safe drinking water (from a protected source with less than 30 minutes for a round trip to fetch water), using good cooking fuel (no wood, coconut shells, or charcoal), having a private toilet, number of rooms and floor area.

Table 2 reports variables related to distance and accessibility of essential services in rural areas. Both surveys include a community questionnaire, 1996 for urban and rural communities and 2009/10 only for rural areas. In both cases they include questions about the time to get to the nearest school, market, town, etc. Some of these questions overlap, while others are similar. Particularly the distance to the nearest road was only asked in 1996. In 1996 the questions were answered in four categories (<30 min, 30-60 min, 60-120 min, >120 min), while in 2009/10 continuous answers were given. For comparability we simply included variables in the table that indicate whether the time to the nearest service is less than 60 minutes.

Table 1: Welfare indicators included in 1996 and 2009/2010 household surveys

	PNGHS 1996			HIES 2009/10		
	Mean	Std. Err.	N	Mean	Std. Err.	N
Age	21.81	0.328	8675	22.25	0.189	22718
Going to school (for school-aged children)	0.449	0.032	1836	0.568	0.025	4004
How long to get to school (minutes)	38.51	3.845	1391	27.28	2.581	5282
Ever been to school (if at least 12 years old)	0.627	0.031	5740	0.744	0.015	15478
Literacy (if at least 12 years old)	0.518	0.036	5740	0.528	0.016	15505
Short of Food	0.219	0.036	1396	-	-	-
Electricity	0.104	0.026	1396	0.134	0.012	4076
Access to safe drinking water	0.365	0.049	1396	0.409	0.027	3987
Good cooking fuel	0.119	0.026	1396	0.111	0.007	4076
Have own toilet	0.815	0.034	1392	0.803	0.016	3683
Household Size	6.014	0.146	1392	5.146	0.067	4081
Number of rooms	3.035	0.108	1392	2.370	0.049	4076
Floor Area	40.87	1.746	1392	47.23	1.507	4072
Headcount Poverty*	0.377	0.025		0.399	0.020	

* Values from technical report on PNG poverty profile by John Gibson 2012.

Table 2: Distance and accessibility indicators for rural communities in 1996 and 2009/2010 household surveys

	PNGHS 1996			HIES 2009/10		
	Mean	Std. Err.	N	Mean	Std. Err.	N
Distance to nearest town (km)				169.83	30.60	105
Nearest town can be reached by road all year round				0.717	0.042	106
Road	0.725	0.063	73			
Community School	0.714	0.065	73	0.703	0.047	106
High School	0.369	0.065	73	0.362	0.048	106
Aid Post	0.715	0.070	54	0.483	0.051	106
Health Center	0.509	0.067	73	0.483	0.049	106
Child Health and Nursing Service (CHNS)	0.696	0.065	73	0.397	0.048	106
Community Health Worker				0.495	0.050	106
Clinic				0.370	0.048	106
Town or Government Station	0.449	0.066	73			
Air Strip	0.344	0.063	73			
Port	0.121	0.048	65	0.166	0.034	106
Phone	0.500	0.067	73			
Internet Service				0.192	0.037	106
Postal Service	0.420	0.066	73	0.235	0.040	106
Daily Market				0.573	0.048	106
Weekly Market				0.643	0.048	106
Store				0.659	0.047	106
Banking Service				0.224	0.040	106
Police Station				0.439	0.049	106
Public Transport				0.573	0.050	106

Accessibility indicators express the fraction of respondents that can reach nearest service in less than 60 minutes.

2. Administrative regions

Spatially explicit data at the census unit level are essential for the current analysis of the impact of road infrastructure as they add locations to the survey data that were described in the preceding part of the report. Data on administrative regions at higher aggregation levels may be useful to describe the context in which local developments take place. Spatial delineations of local and more aggregate administrative regions have been collected from various sources as is described below. Figure 1 shows the included regional divisions and the selected Highlands region study area.

2.1. Census units

The 2009/2010 household income and expenditure survey (HIES) that captures local socio-economic conditions contains reference to the 2000 census units in which the interviewed households live. By including a spatial representation of the census units we are able to locate these households and describe their proximity to road infrastructure or other spatial characteristics. This step is essential as the location information included with survey is too coarse (latitude/longitude coordinates are described with a 0.1 degree resolution which implies a spatial accuracy of approximately 50 km).

A spatial representation of the census units (describing point locations in ArcGIS shapefiles) was obtained from the Pacific Catastrophe Risk Assessment and Financing Initiative (PCRAFI) portal (http://pcrafi.sopac.org/layers/geonode:pg_census_unit#more). These point data were used to georeference the survey data and can be used to construct, amongst other, the distance to the nearest road. The census unit data from the PCRAFI portal also contains reference to the population totals for 2000 (based on the census) and estimates for 2008 and 2010.

2.2. Provinces, districts and lower level government

Shapefiles describing the borders of various regional administration levels have been collected from the Humanitarian Response portal (www.humanitarianresponse.info/en/operations/papua-new-guinea/datasets). The initial shapefiles originate from the National Statistics Office of Papua New Guinea. These spatial delineations are useful to store region-specific statistical data on themes such as economic development, demography et cetera. Many socio-economic attributes from the 2000 census have been added to the different administration levels providing a wealth of information related to population number, demographic characteristics, household composition, income etc. The data is consistent across the different administrative levels, meaning that population counts in the provinces add up to those of the underlying districts and lower level government regions.

The local level government regions were used as the finest spatial level in the poverty mapping of Papua New Guinea performed by Gibson and others (2005). The tabular data of this analysis is available from the SEDAC portal to Environmental and Socioeconomic Data (http://sedac.ciesin.columbia.edu/data/set/povmap-poverty-food-security-case-studies/data-download). This portal is hosted by the Center for International Earth Science Information Network (CIESIN) at Columbia University as part of NASA's Earth Observing System Data and Information System (EOSDIS). Using additional data obtained from two researchers who participated in the poverty mapping study (Guarav Datt and Bryant Allen) we were able to reconstruct their maps depicting regional poverty levels in 2000 (Figure 2).

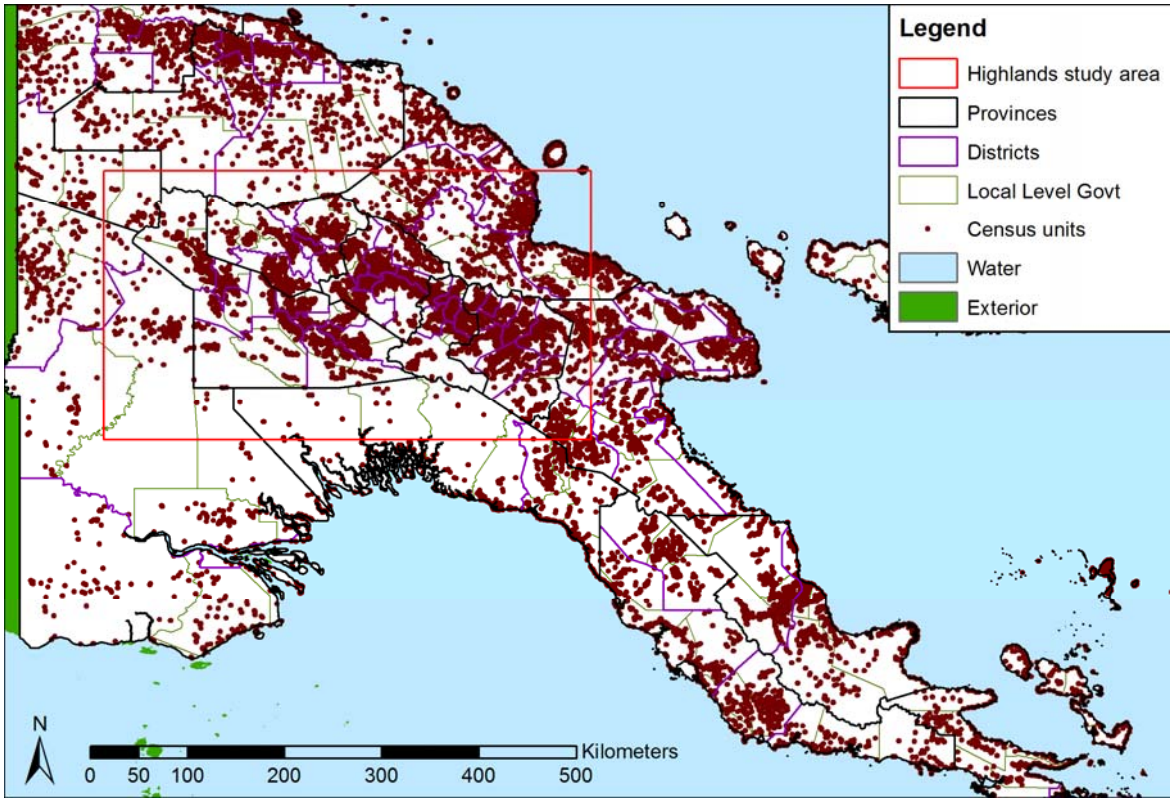

Figure 1 Regional divisions and highlands region study area (source: www.humanitarianresponse.info).

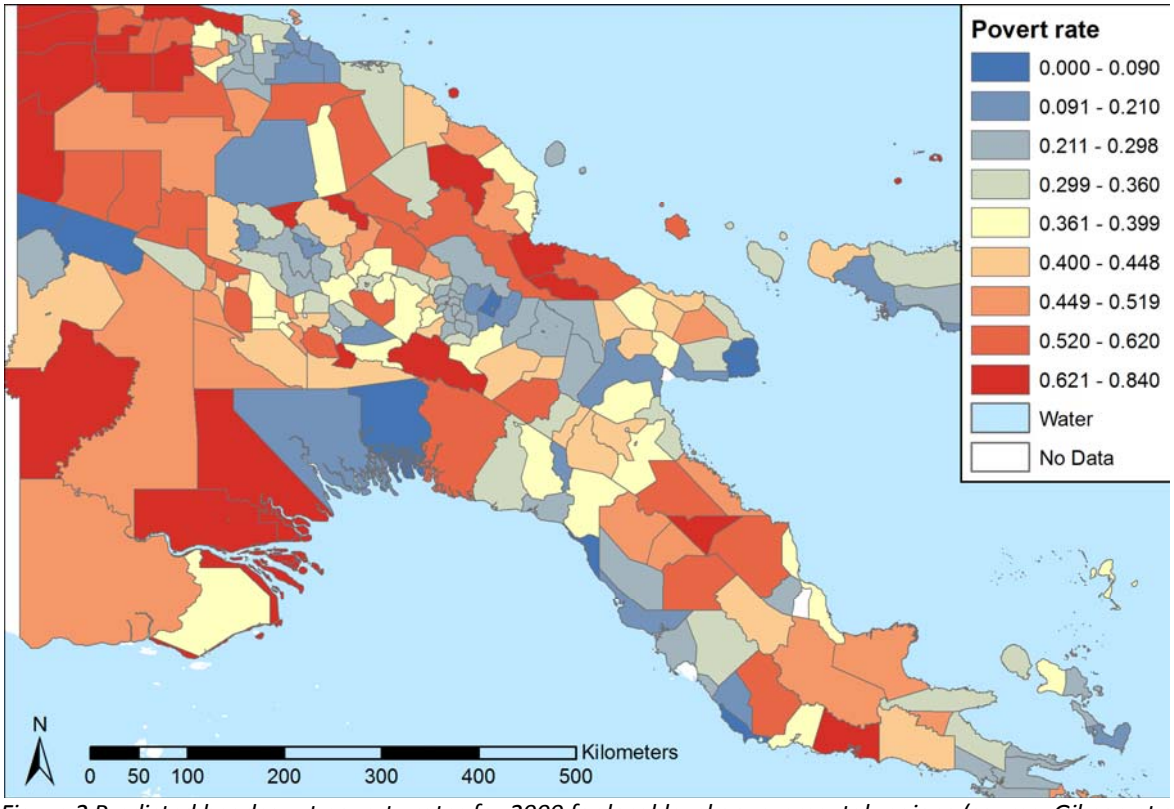

Figure 2 Predicted headcount poverty rates for 2000 for local level governmental regions (source: Gibson et al., 2005).

3. Population density data

Several globally available spatially explicit datasets on population density have been collected to help characterise spatial patterns in population. These datasets typically downscale regional-level population statistics to the local (grid cell) level following simple assumptions about their spatial distribution. The collected data have been developed by different academic institutions and are freely available to the international community.

3.1. Gridded Population of the World

The Gridded Population of the World (GPW) is a spatially disaggregated population data set with a 2.5 arc-minute resolution, which is equivalent to approximately 5x5 km in Papua New Guinea. The data is provided by the SEDAC Portal. Included here is version 3 of this data set with population counts (number of inhabitants per grid cell) and population density (inhabitants per km^2) for the years 1990, 1995 and 2000. More detailed information on this data set is provided elsewhere (e.g. Balk et al., 2006, or the download page at: http://sedac.ciesin.columbia.edu/data/collection/gpw-v3). The data is created by distributing regional population counts homogenously over all grid cells in a region. In the case of Papua New Guinea this implies that grid cells within a province receive the same hypothetical value. This makes the data only useful for general illustration purposes.

The GPW data is also available for population projections relating to 2005, 2010, and 2015 (observation data for these years was not available when the datasets was constructed in 2004). In addition, a preliminary release of version 4 of this dataset has recently become available (http://www.ciesin.columbia.edu/data/gpw-v4/). The new dataset describes the number of persons and population density per 30 arc-second grid cell for 2010.

3.2. Global Rural Urban Mapping Project (GRUMP)

Part of the limitations of the GPW data are overcome in the Global Rural Urban Mapping Project (GRUMP) dataset that builds on the GPW approach, but incorporates light at night data from satellite to differentiate urban and rural areas in the spatial reallocation of population for each census block). This data has a higher resolution (30 arc-seconds, or circa 1x1km at the equator) than the GPW data and describes population patterns with more spatial detail, highlighting population concentrations in larger communities and characterising the surrounding, rural areas with lower values. Yet, it still characterises large areas with uniform hypothetical values that do not match the actual dispersion of population. The data is available for the same years as the GPW data (1990, 1995, 2000) using the same regional-specific input data. See Balk et al. (2006) for more details (or consult the download page at: http://sedac.ciesin.columbia.edu/data/set/grump-v1-population-count/data-download.

A comparable data set is also available from the LandScan Global Population database that is updated annually. This dataset applies a larger set of ancillary data, such as land cover, transport network and topographic data to re-distribute census data in a gridded format (see Bhaduri et al., 2007, or: http://web.ornl.gov/sci/landscan/). A limitation of this data set, however, is that it is not freely available for download and that its assumptions related to input data and spatial distribution methods are not fully disclosed (Gaughan et al., 2013). Instead, the Worldpop data has been selected that has become available recently and that is described below.

3.3. Worldpop

The WorldPop project aims to provide an open access archive of spatial demographic datasets for Central and South America, Africa and Asia (http://www.worldpop.org.uk/). It was initiated in 2013 to combine earlier mapping efforts for these regions. WorldPop relies on detailed (30 metres resolution) Landsat Enhanced Thematic Mapper (ETM) satellite imagery and additional sources to downscale regional population counts. More extensive descriptions of the applied methodology and incorporated base data are available in several scientific paper (Gaughan et al., 2013; and Stevens et

al., 2015). For Papua New Guinea data population counts are available for highly detailed grid cells with a resolution of 0.0008333 decimal degrees (circa 100x100m) for the years 2010 and 2015. A first visual inspection reveals a highly disaggregate pattern with much more detail than the GRUMP data. It is unclear, however, how realistic these simulated patterns are.

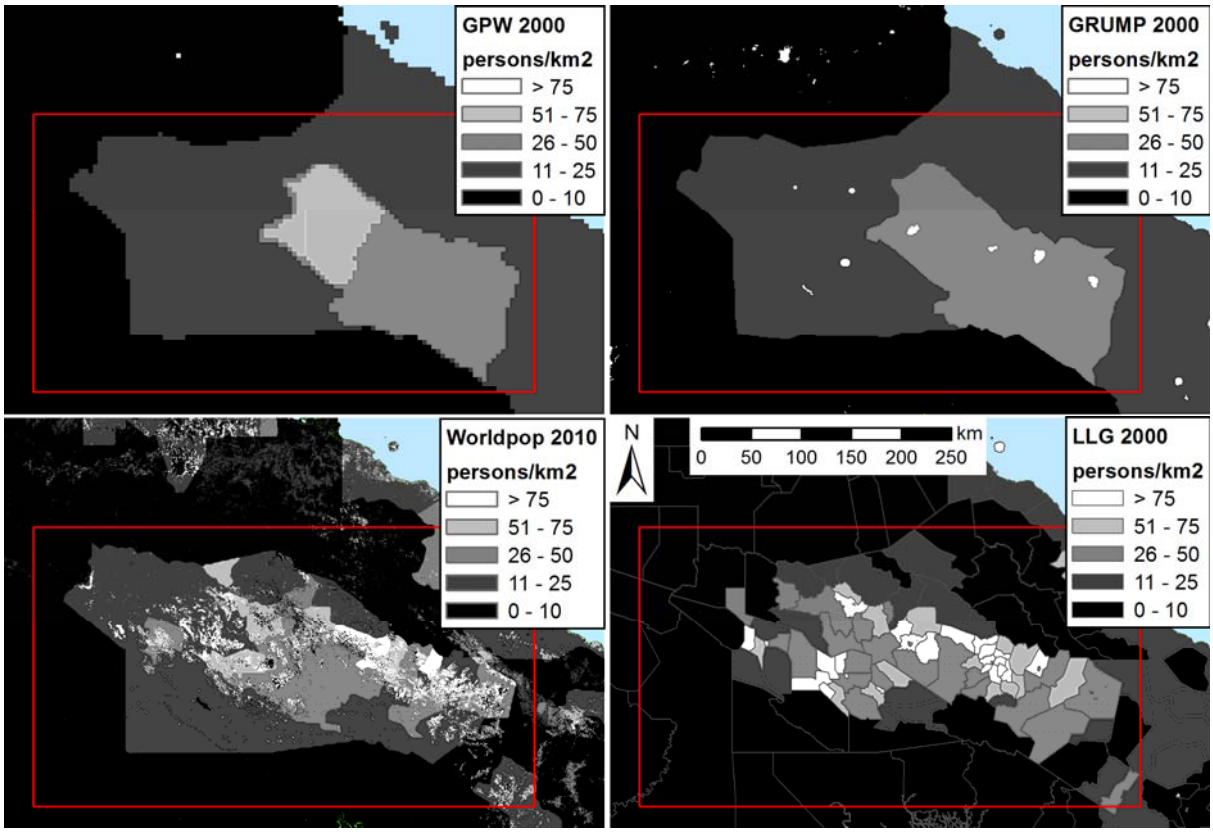

Figure 3 Different gridded population density datasets compared with census 2000 data describing population per Lower Level Government (LLG) level (at bottom at right). Note that grid cell size differs per data set. Census data derived from the Humanitarian Response portal (www.humanitarianresponse.info).

4. Road network

4.1. National Mapping Bureau (NMB)

The National Mapping Bureau of Papua New Guinea has created a comprehensive spatial dataset of the 2000 road network. This data set was obtained from the humanitarian response portal (www.humanitarianresponse.info/en/operations/papua-new-guinea/datasets). The included data contains attribute information for different road segments describing name (where present), surface (sealed, gravel, dirt), road type (highway, major, minor, track)and condition (good, average, bad and F).

Table 1 lists the frequency of occurrence of the different types of roads and their conditions in the highlands study area. These attributes are of particular relevance to our study and need to be checked with local partners to ensure their validity. Figure 4 and Figure 5 show the type and condition of the road segments overlaid on an elevation map that is further described in Section 5. A further discussion on obtaining road quality data from remote sensing is provided in Section 9.

Table 1 Cross comparison of road type (rows) and road condition (columns) of the road segments (in km) in the NMB 2000 dataset for the highlands study area

Type \ Condition	Good	Average	Bad	F	Unknown	Total
Highway	28	13	10			51
Major	17	36	54		1	108
Minor	27	153	309		8	497
Track	2	10	29			41
Unknown				43	55	98
Total	74	212	402	43	64	795

4.2. Open street map (OSM)

An additional source for road network data is the Open Street Map dataset that is maintained by a community of voluntary mapmakers (www.openstreetmap.org). This open source data set contains many types of data sets (roads, waterways, places, points of interest etc.) and is updated continuously. An extract relating to Papua New Guinea was created from this data set on May 20, 2015 through the BBBike portal (http://BBBike.org). The road network in this extract includes attribute information on, for example, road types and road names in English(where known). Figure 6 shows the different types of roads that are distinguished in this network for the highlands study area. The OSM dataset contains less roads than the NMB data set (e.g. in the Enga and Eastern Highlands provinces), but does include some additional links (e.g. towards the Gulf province) that may not have been present in 2000. As the NMB data set seems more comprehensive and describes road quality this seems to be a more logic starting point for our analysis, but the OSM data set may be used to fill in possibly missing road segments for studies related to more recent data.

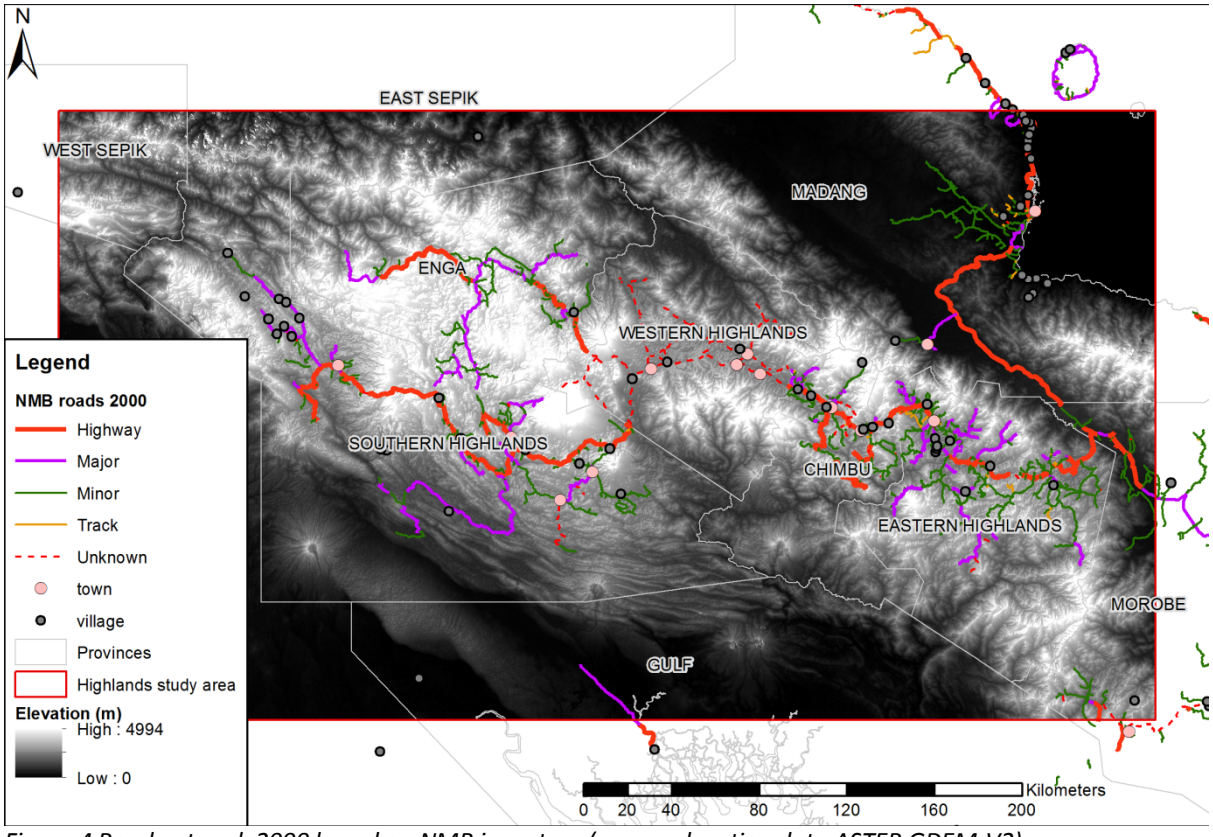

Figure 4 Road network 2000 based on NMB inventory (source elevation data ASTER GDEM-V2).

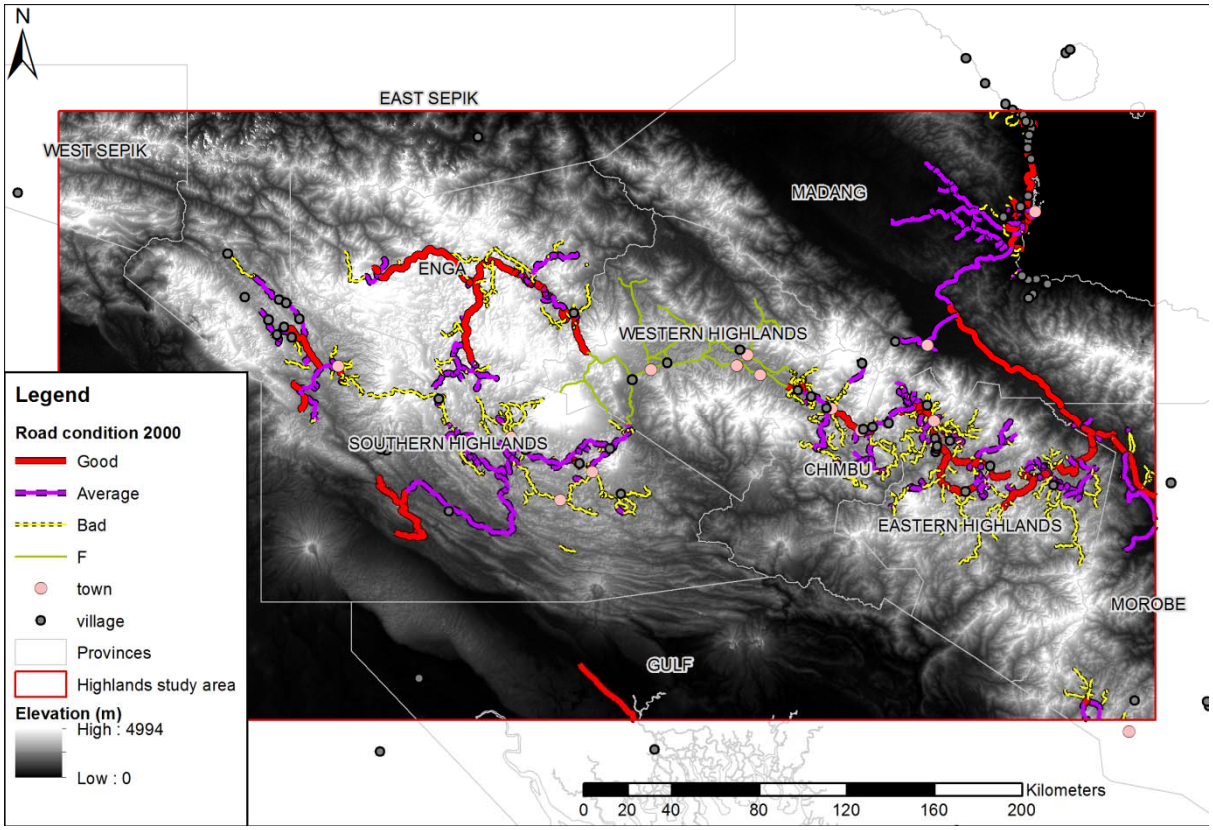

Figure 5 Road condition 2000 in study area based on NMB inventory (source elevation data ASTER GDEM-V2).

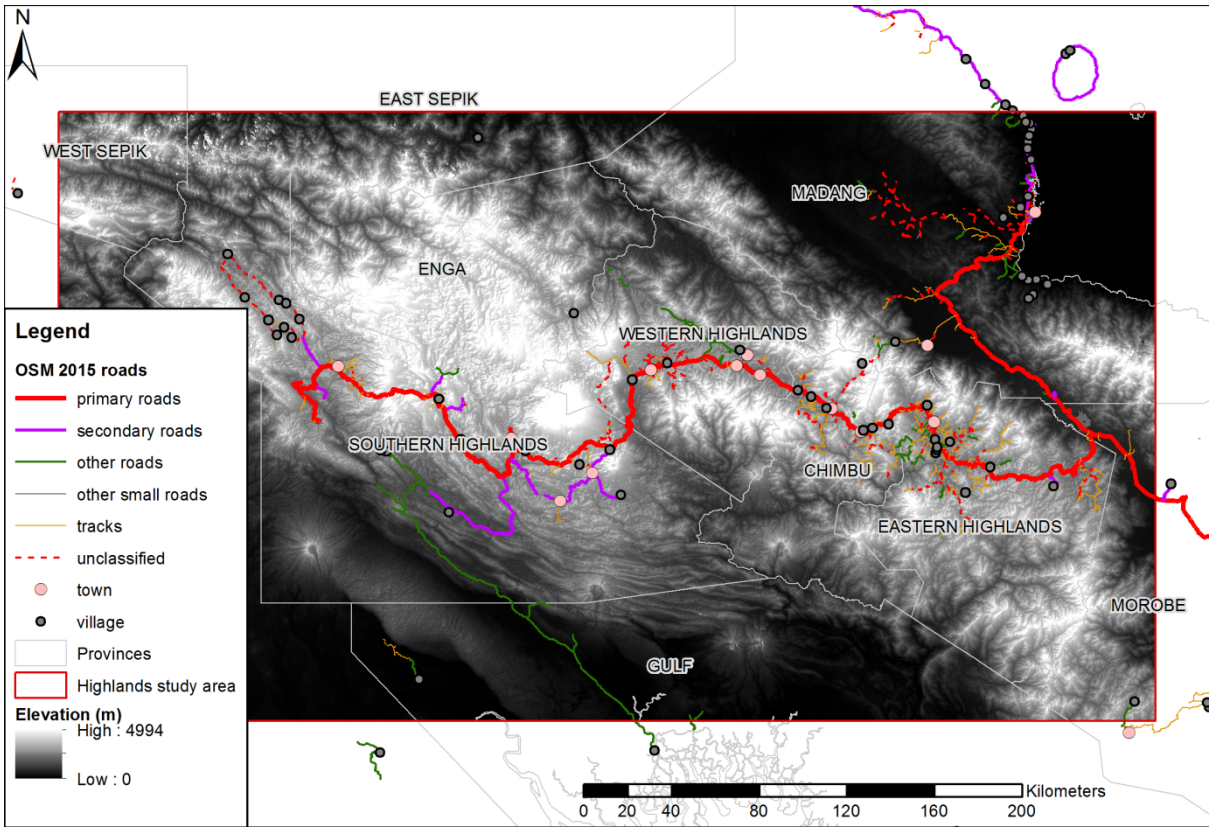

Figure 6 Road network 2015 based on OSM data (source elevation data ASTER GDEM-V2).

4.3. Accessibility

A proper spatial representation of the road network is a first step in developing accessibility-related indicators that express the importance of roads for local communities. A second step is to define in which way roads matter to the local population. Do they, for example, require access to major towns in the region for the provision of specific goods and services? Or is it important that they are within walking distance of a major road to be able to sell produce? Or is the road-based distance to the nearest medical facility an important indicator of their health status? For these and other benefits of accessibility, spatially-explicit indicators can be calculated using the road network and possibly also other characteristics of the terrain (e.g. slope or land use as factors that impede transport).

Figure 7 provides an example of an accessibility indicator depicting the travel time to the nearest major city (in this case Lae as access point to the coastal region) in hours. Travel times are calculated assuming motorised transport over roads and rivers and land-use specific walking speeds over other terrain. This indicator is provided as Global Accessibility Map by the Joint Research Centre of the European Commission and makes use of, amongst others, a global road network (vmap0) and a spatial inventory of navigable rivers (for more information on the applied methodology and incorporated data sets see: http://forobs.jrc.ec.europa.eu/products/gam/sources.php). While this indicator does not seem appropriate for the current study because of its focus on long-distance travelling and the inclusion of an incomplete road network, elements of the methodology could be applied in calculating accessibility indicators for Papua New Guinea. Section IV discusses our own attempts to calculate road network based accessibility measures.

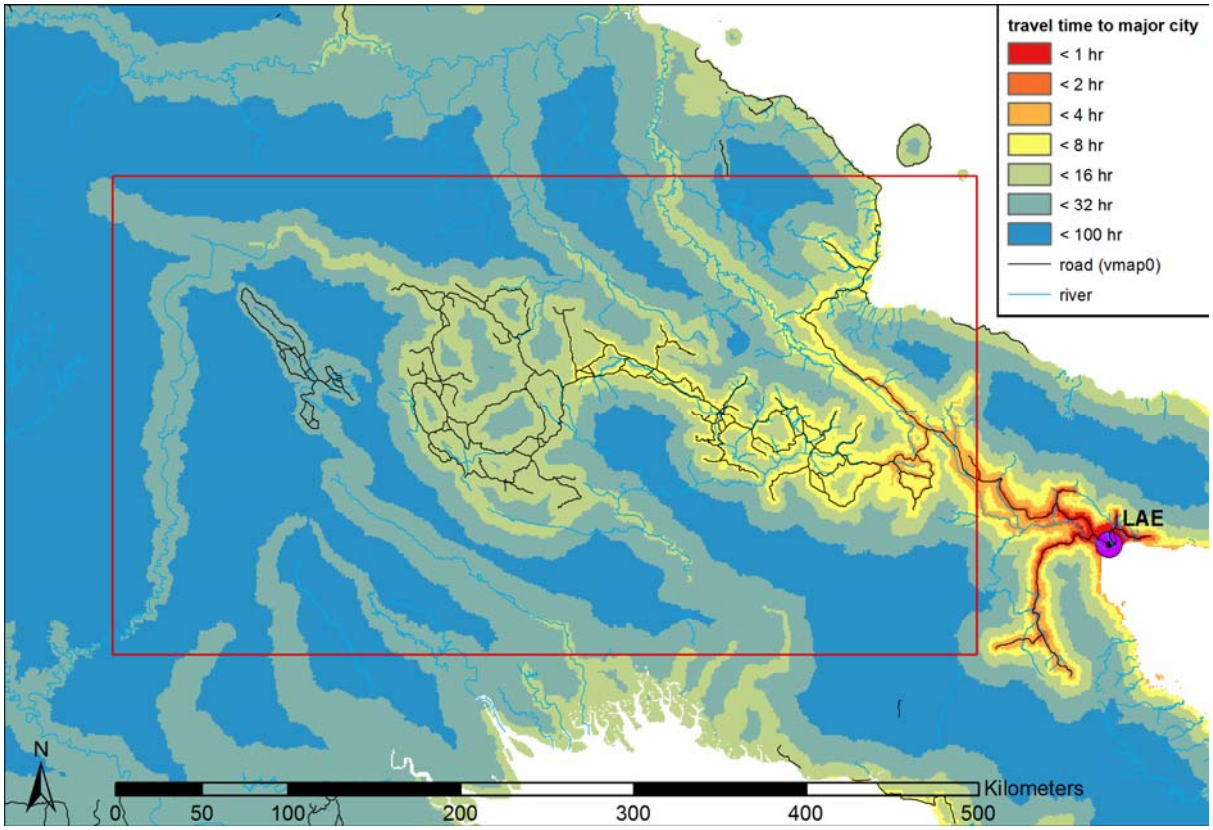

Figure 7 Example of an accessibilty related indicator: travel time to nearest major city in hours (source: EC-JRC). The figure also depicts the global (vmap0) road network used in these calculations.

5. Elevation

Elevation data was obtained from the Global Digital Elevation Model Version 2 (GDEM V2). This data is collected with the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) sensor as part of a joint American-Japanese remote sensing mission. Due to their global availability

and relatively high resolution and accuracy ASTER-GDEM data are often used in spatial analyses when more detailed local data sources are lacking. The initial data has a resolution of 30x30 metres and an average absolute vertical accuracy of 0.2 metre (Meyer, 2011). In mountainous forested areas (such as the Papuan Highlands) the data has an offset (consistent error) of circa +7 metres with a standard deviation of around 13 metres (ibid.). Seeing the large range of elevation values in the study area (from 0 to 4500 metres) this minor error does not hamper its inclusion in our analysis.

In total 24 individual map sheets were downloaded from the Earth Remote Sensing Data Analysis Center of Japan (ERSDAC, see: http://gdem.ersdac.jspacesystems.or.jp/download.jsp). These sheets were first clipped to the match the boundaries of the Highlands study area and then mosaicked into one data set. Unlikely values in the resulting composite (elevations below sea level and above the highest peak of 4510 metres) were replaced by average elevation values of neighbourhood grid cells in iterative process. During this process the elevation data is aggregated to a 90x90 metres resolution to limit files size and facilitate faster representation. This resolution is deemed sufficient for further analysis.

The figures in the preceding section show the resulting data set a background for the road network of the region. The processing of the data was performed with scripts that allow the repetition of the data treatment steps at a higher resolution if required.

6. Luminosity

Luminosity or light at night (LAN) data is used extensively in attempts to map socio-economic features. Early examples include the analysis of urban energy consumption (Welch, 1980) and the mapping of urban areas (Imhoff et al., 1997) and population densities around the globe (Elvidge et al, 1997; Dobson et al., 2000). Various studies have indicated a strong correlation between total national emittance of light, population density and especially economic activity (Doll et al., 2000; Elvidge et al., 2001). At the national scale the latter relationship seems to be stronger due to the direct relation between prosperity and usage of electricity for, for example, the outdoor lighting of buildings, towns and roads. More recent studies focussed on the mapping of economic activity at a more regional and local scale (e.g. Doll at al., 2006; Sutton et al., 2007; Ghosh et al., 2010; Levin and Duke, 2012) or on analysing the differences between light emittance and population density to study phenomena such as poverty (Elvidge et al., 2009), access to electricity in rural areas in developing countries (Doll and Pachauri, 2010) and the size of the informal economy in Mexican states (Ghosh et al., 2009).

Elvidge et al (2001) found that population densities in rural areas are difficult to map with LAN data: even in the prosperous United States rural towns are only consistently detected when they have more than 200 inhabitants. They expect these detection limits to be substantially higher in less developed areas of the world. This assumption was later confirmed by Doll and Pachauri (2010) when they compared the percentage of illuminated pixels with local population density for different continents. Whereas in Europe and the United States 90% of the locations with a density of 50 or more persons per km^2 was found to be lit, this share of lit pixels dropped to 10% in developing countries in Asia. Even at densities of 250 persons per km^2, only 50% of the pixels was lit in this region. While this discrepancy may be partly caused by the poor quality of the GRUMP data that are used to describe population density (as is also apparent from Figure 3 in this report), it also indicates that rural populations in developing countries produce too little light to be consistently detected. According to the study of Doll and Pachauri 99.1% of the rural population of Papua New Guinea was unlit in 2000.

Initially, LAN data was obtained with the Operational Linescan System (OLS) on board of satellites that were launched as part of the United States Defense Meteorological Satellite Program (DMSP). For the period 1992-2013 these images are available from a range of different satellites (see:

http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html). Recently, night time light information has become available through the Visible Infrared Imaging Radiometer Suite (VIIRS) instrument. This data is expected to provide more information on low population and economic density areas than the stable lights data generated by the DMSP-OLS system (Chen and Nordhaus, 2015). VIIRS data from 2014 onwards is available from the Earth Observation Group, NOAA National Geophysical Data Center (http://ngdc.noaa.gov/eog/viirs/download_monthly.html). The images express observed luminosity values averaged per month based on the number of cloud free days. This approach corrects for the fact that some locations will have lower luminosity because they are more often covered by clouds. Obviously these images will not capture luminosity for those locations that were covered by clouds on all observation days.

6.1. Comparing different luminosity data sources with population density

To test the relevance of LAN data socio-economic research in Papua New Guinea as suggested in the Terms of Reference for this project we obtained OLS and VIIRS data sets and compared these with available data on population density. In order to provide the best possible match, OLS data was collected for the most recently available year (2013), while the oldest possible year was collected for the VIIRS data[1] (2014). OLS data is available in two forms: an average luminosity value per year and a so-called stable lights version that excludes erratic light occurrences (e.g. related to wild fires, gas flaring etc.) from calculating those averages. We selected the stable lights version as that excludes the back ground noise that may obscure actual settlements. For the VIIRS data a temporal composite was created by combining the images of different months (weighing the monthly images by the number of cloud free days to give more weight to more reliable estimates). This operation resulted in an average observed luminosity per cloud free day for the year 2014 (Figure 9). The VIIRS images have a 15 arc-seconds resolution (circa 500x500m), while the OLS data have a resolution of 30 arc-seconds (circa 1x1km). The finer resolution and different sensor allow the depiction of more detail in the observed amount of light at night. This increased resolution comes at a price however, as the images also show more noise and oversaturation (dispersion of light into neighbouring grid cells).

Visual inspection of the LAN data learns that major centres such as Port Moresby and Mount Hagen are well captured in the images. Smaller settlements have much lower luminosity values, but can still be observed. The most striking features, however, are mining and oil production sites. In the highlands the Porgera and Ok Tedi mines as well as the Kutubu-Iagifu, Agogo and Hides oil and gas production sites with their related infrastructure stand out as high luminosity areas. To assess the potential of LAN-data in mapping the poverty and welfare in the communities of the Highlands we have excluded these mining sites from our further analysis as they seem to bear little relation with the distribution of settlements (represented as census unit sites) across the region.

In order to assess the potential of luminosity data for capturing concentrations of people, a comparison was made of population density and luminosity values at the level of Local Level Government (LLG) regions: the lowest spatial aggregation level at which population data was available. The analysis was performed for the mainland of Papua New Guinea. The area covers about 432,000 km^2 and is comprised of 249 LLG regions that range in size from 1 to 23,000 km^2. Population density for 2000 (expressed in persons/km^2) and average luminosity values per area unit[2] were

[1] Standard average luminosity data was used instead of the version that includes the stray-light corrected data as this set mainly has added values toward the poles and is of reduced quality (for more information on this issue, see http://ngdc.noaa.gov/eog/viirs/download_monthly.html and Mills et al., 2013).

[2] VIIRS luminosity values are expressed in nanoWatts/cm^2/sr and multiplied by 1E9 to obtain readable figures. OLS luminosity is described with dimensionless digital numbers ranging from 0 to 63. In our analysis we sum all luminosity values for a region and divide the obtained sums by the size of the area (in square decimal degrees). We focus on densities of people and luminosity (rather than aggregate, total values) as the regions differ strongly in size. Many of the smaller

found to have a fairly strong and positive linear correlation (Figure 9). This correlation is stronger for VIIRS luminosity (R^2 = 0.43) than for OLS luminosity (R^2 = 0.34) indicating that the former data set is indeed better equipped to map population distribution. The figure also suggests that this relation is stronger for higher population densities. This suggestion is confirmed when we run separate regressions explaining VIIRS luminosity from population density for observations with density values below 25 persons/km^2 (R^2 = 0.01) and above that value (R^2 = 0.52). The variation in luminosity for low population densities will thus be caused by other factors and may link to, for example, noise in measurements (e.g. caused by wildfires or measurement errors), proximity to larger towns (whose light may pollute neighbouring rural areas) or the presence of specific facilities (airports, harbours etc.). Figure 8 and 9 also indicate that both sources of luminosity data capture fairly similar patterns. That is also reflected in the strong degree of correlation in luminosity values on both data sets (R^2 = 0.72).

A comparison with the economic performance of these regions is not directly possible as most economic statistics are only available at the national level. An attempt was made to downscale this data to the grid cell level using the method that is used to create the G-econ database on gridded output (Nordhaus et al., 2006). This study is not included here as its main input is gridded population density data that is not providing a trustworthy representation of reality (see Section 3). A brief report describing this effort is available for download at the VU-SPINlab website: www.feweb.vu.nl/gis/publications/docs/Using_LAN_data_as_proxy_for_economic_activity.pdf

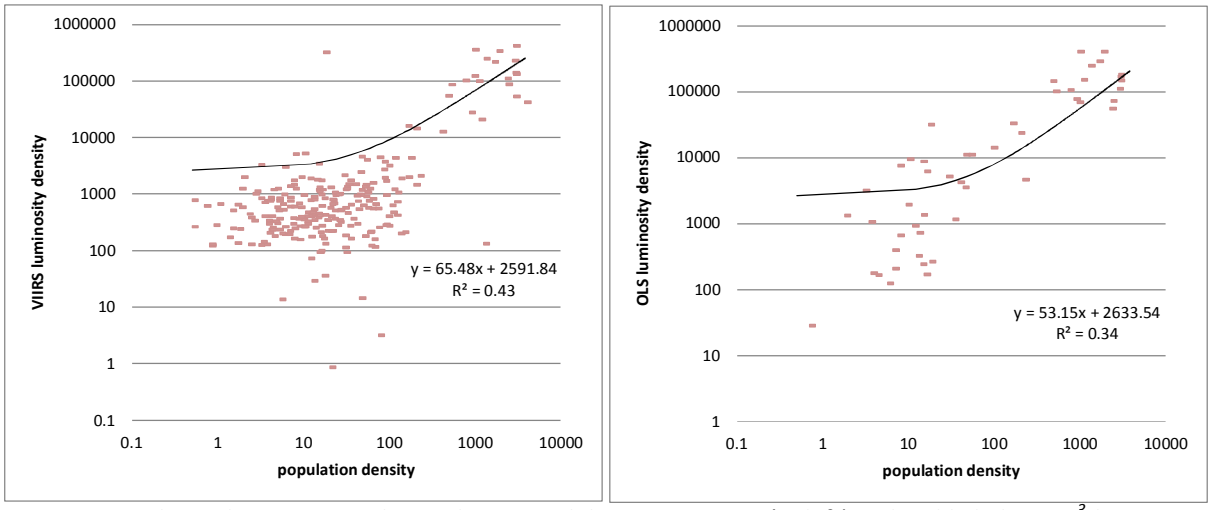

Figure 8 Correlation between population density and densities in VIIRS (at left) and stable lights OLS[3] luminosity at the level of Lower Local Government regions. The right-hand side figure only shows the 55 regions with a value higher than 0. The 0-values are not visible because of the logarithmic scale that was applied to map the extreme range of density values. VIIRS-luminosity was observed in all 249 regions.

regions represent densely populated areas that have high luminosity values for similar amounts of people than the much larger sparsely populated regions that are characterized by very low or absent luminosity values.

[3] A similar analysis was performed for the OLS data with average luminosity values that do not exclude background noise. This non treated data set did not show a relation with population density.

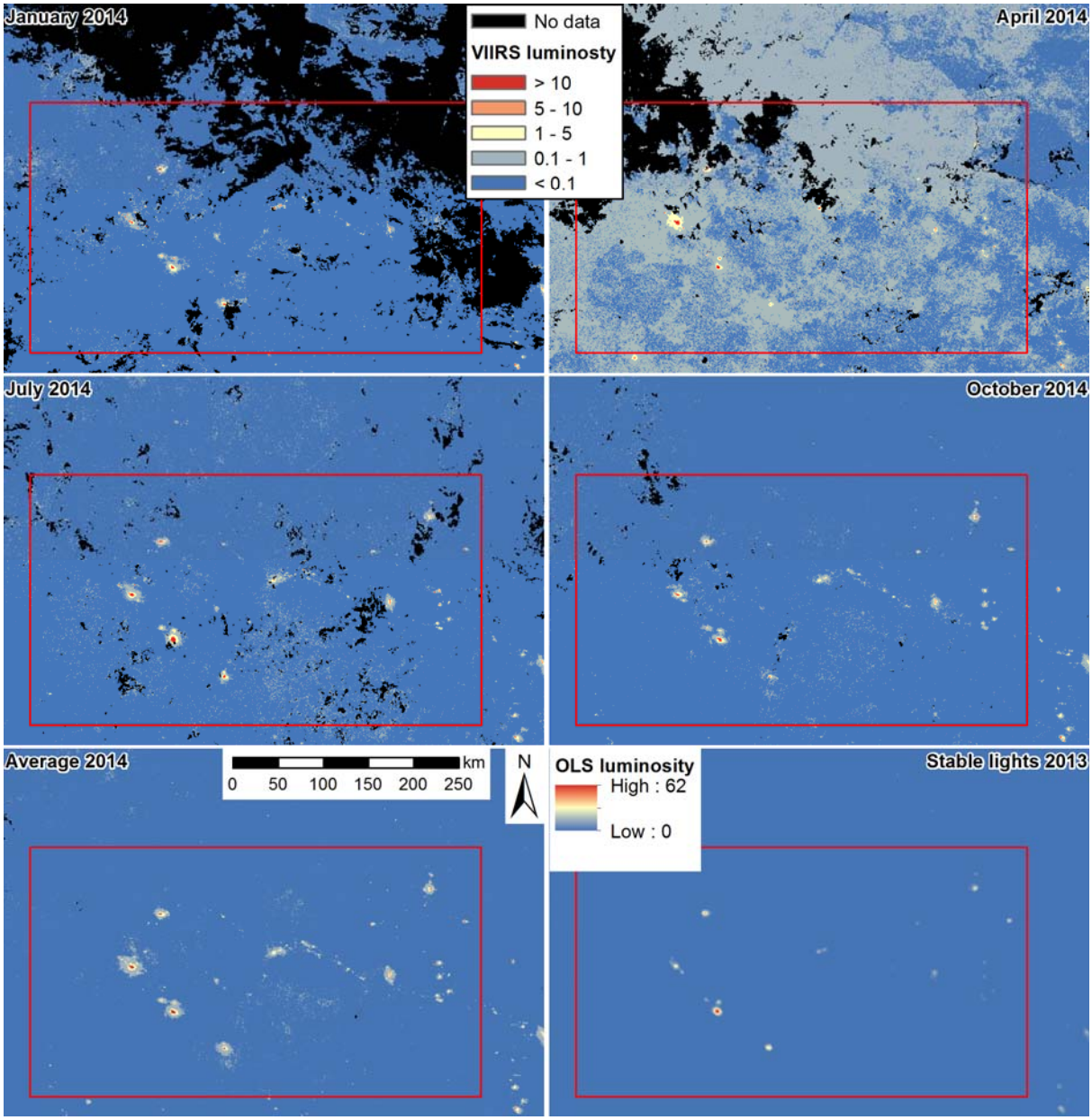

Figure 9 Luminosity data from the VIIRS dataset compared to the OLS stable lights data set (bottom right). VIIRS were collected for four months in 2014 and subsequently averaged to obtain an image representing the study area without no data values (bottom left), Source for both datasets: http://ngdc.noaa.gov.

6.2. Building a time series of luminosity data

Luminosity data is available since 1992 and may be useful to discern trends in development. To test this we built a time series of luminosity data. Here we will briefly highlight the findings for the period 2000-2010 that is of particular relevance for our assessment of the impact of roads on household income.

Figure 10 shows the OLS stable lights data for 2010 and 2000 in the central part of the Highlands region where luminosity values are not influenced by the presence of mines or oil production facilities. The figure shows the original luminosity values as well as values that are corrected for differences in GDP and total amount of light per year. In the latter representation each luminosity unit corresponds to a fixed contribution to GDP. This contribution is calculated following a correction factor that divides the total GDP in a particular year by the total amount of light of that year. This factor thus expresses economic intensity per light unit and assumes that all economic activity is

reflected in luminosity which is of course not true. Ideally we would calculate this factor only for the share of economic activity that is likely to be captured by light (most activities in larger towns, mining etc.). But as long as this share is constant over time this approach helps correcting the differences in overall reflection per year. Table 2 provides an overview of national GDP, total amount of observed light and resulting economic intensity values for the years luminosity data was collected.

Table 2 Key statistics of the collected OLS stable lights data

Year	Satellite	GDP[4] (billion PNG kina)	Change (% / year)	Total luminosity	Economic intensity (million kina / light unit)
1992	F10	6.2	n.a.	23376	0.2642
2000	F14	7.8	3%	43482	0.1783
2000	F15	7.8	n.a.	50048	0.1549
2010	F18	11.5	5%	45683	0.2522
2013	F18	14.5	9%	40074	0.3629

The difference maps included in Figure 10 show that the correction procedure is able to limit the impact of differences in the way light was detected in the two years. The direct comparison of luminosity values suggests strong local increases and decreases in the Western highlands province (bottom-left map in Figure 10), while the adjusted comparison suggests a small overall increase (bottom-right map in Figure 10). Similar observations can be made for the Southern highlands province, while very little seems to have changed in Enga province. This approach may be helpful in trying to assess changes between years, but will still be sensitive to measurement differences that can relate to atmospheric conditions, time at night data was captured et cetera.

The described analysis can be enhanced further. It could, for example, be more meaningful to use regional GDP values for the proposed correction procedure and possibly even correct for the regional differences in the structure of the economy. Another interesting alternative could be to map the population concentrations that are not captured by light: these must be relatively poor. Or we could create aggregate values (totals per region) and compare these with regional statistics. Moreover, the LAN-data time series can be used to map sudden large changes in economic activity, such as the opening of new mines/oil fields. In the collected time series the southernmost oil field was not yet present in 1992 while the other main mines and oil fields are visible in all years. For our current analysis of household poverty status, however, we feel that luminosity data provides no additional value to the detailed household level surveys that we have available.

[4] GDP is reported in constant prices, showing the values in market prices of the base year (1998) to adjust for the impact of inflation. Data was collected from World Economic Outlook Database of the International Monetary Fund (http://www.imf.org/external/pubs/ft/weo/2015/01/weodata/index.aspx). The original source of the GDP data is the Papua New Guinea National Statistical Office and Ministry of Finance.

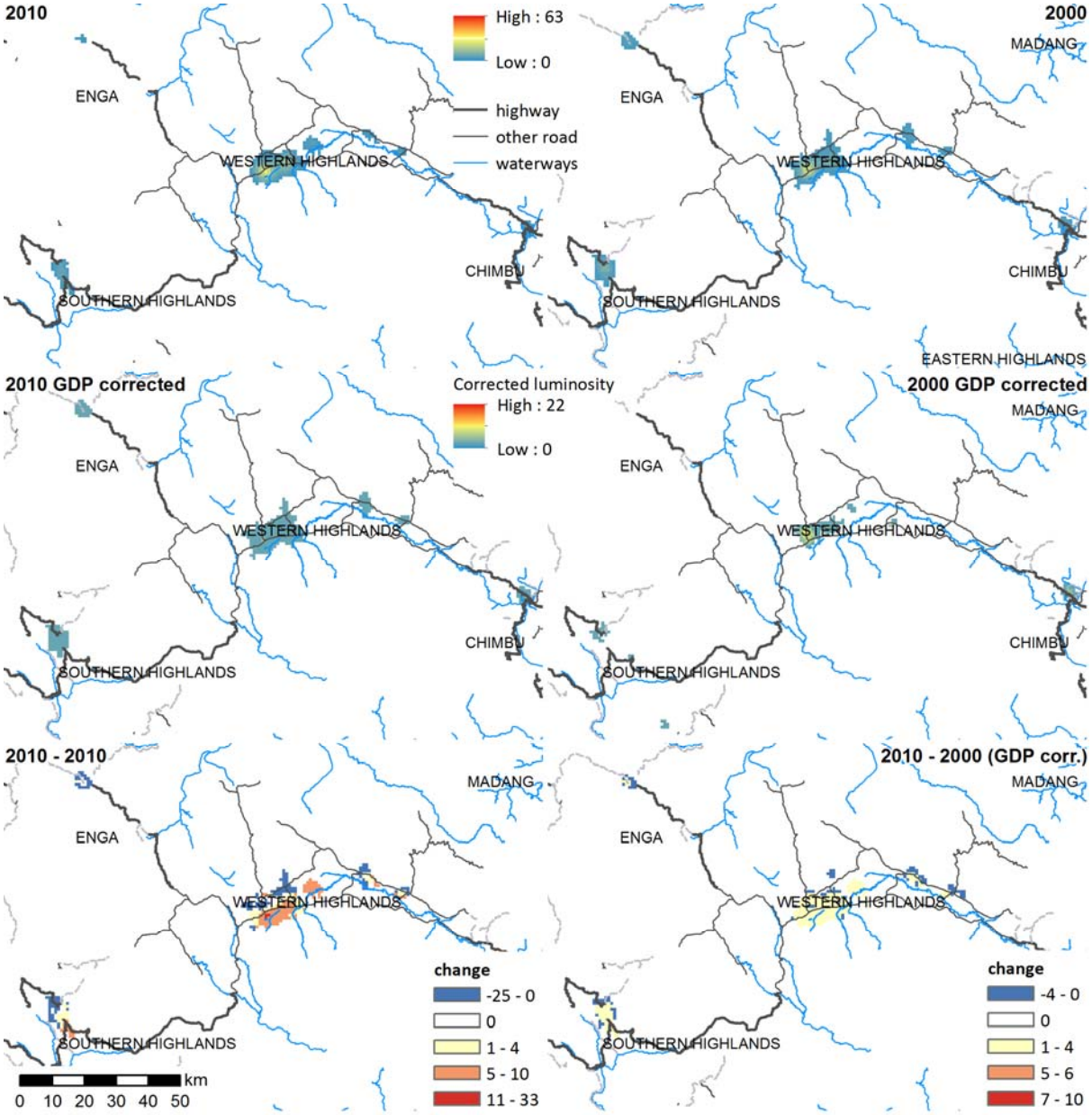

Figure 10 Comparison of OLS stable lights data for 2010 and 2000 in part of the Highlands region. The top two maps show the original luminosity values, the middle two maps show values that are corrected for differences in GDP and total amount of light per year and the bottom two maps compare the two years.

7. Additional thematic data sets

Spatially explicit data sets describing specific socio-economic or biophysical themes can be applied to enhance our analysis of household income. They can help sketch the local conditions in which the surveyed households operate (e.g. describing amount of rainfall or soil conditions that may favour agricultural production). Or they can be used to characterise the regional environment in which household operate and thus be helpful to upscale local observations.

7.1. Land cover

Land cover data sets can helpful in characterise locations by describing the presence of urbanised areas or agricultural land use. An overview of globally available raster-based land cover maps is provided in Table 3. The most widely applied data sets are: MODIS because of its high temporal resolution that potentially allows the detection of changes in specific short time period; and GlobCover and CCI-LC because of their finer spatial resolution of around 300m. The recently

released GlobeLand30 datasets have an even finer resolution (30m!) but currently only the 2010 version is available for download.

Table 3 Overview of Global land cover data sets (in raster format).

Product	Source	Accuracy	Geographical coverage	Spatial resolution	Temporal resolution	Thematic resolution	Download link
GLC 2000	Based on SPOT 4 satellite imagery	66%- 69%	Regional and Global (aggregated dataset)	1 km	Only available for 2000	Land cover (FAO-LCCS)	http://bioval.jrc.ec.europa.eu/products/glc2000/products.php
MODIS Land Cover	Based on MODIS satellite imagery	2005: 75%	Global (mosaics and aggregated dataset)	500m (mosaics) or 5' and 0.5° (aggregated global dataset)	Every year between 2001-2012	Land cover (IGBP)	ftp://glcf.umd.edu/glcf/Global_LNDCVR/UMD_TILES/Version_5.1/
GlobCover	Based on MERIS satellite imagery	2005: 73% 2009: 68%	Global (aggregated dataset)	300m	2005 and 2009	Land cover (FAO-LCCS)	http://due.esrin.esa.int/globcover/
CCI-LC	Based on MERIS and SPOT-Vegetation satellite imagery	2008-2012: 74%	Global (aggregated dataset)	300m	1998-2002, 2003-2007 and 2008-2012	Land cover (FAO-LCCS)	http://maps.elie.ucl.ac.be/CCI/viewer/download.php
Global Land Survey (and derived products Landsat Tree Cover Continuous Fields and Landsat Forest Cover Change)	Satellite imagery collected from Landsat sensors	-	Global (mosaics)	30m	1975, 1990, 2000, 2005 (LTCCF and LFCC only available for 2000 and 2005)	HR satellite imagery, Tree cover, Forest cover change	GLS: http://glcf.umd.edu/data/gls/ TC: http://glcf.umd.edu/data/landsatTreecover/ FCC: http://glcf.umd.edu/data/landsatFCC/
FROM-GLC 30m	Based on Landsat TM/ETM+ satellite imagery	64%-66%	Global (mosaics)	30m	Only available for 2006	Land cover (compatible with IGBP and FAO-LCCS)	http://data.ess.tsinghua.edu.cn/
GlobeLand30	Based on Landsat TM/ETM, Chinese HJ-1 and auxiliary data	80%	Global	30m	Only 2010 data downloadable, 2000 data have been published	Land cover in 10 classes	http://www.globallandcover.com/

By way of example we briefly describe the often applied MODIS Land Cover here. Other land cover data sets are constructed in a similar way. MODIS offers a series of annual global land cover datasets derived from Moderate Resolution Imaging Spectroradiometer (MODIS) observations by NASA's Terra and Aqua spacecraft. This project is part of NASA's Advancing Collaborative Connections for Earth System Science (ACCESS) program, with the collaboration of the Joint Global Change Research Institute and the Global Land Cover Facility of the University of Maryland. The land cover maps are

generated using a supervised artificial neural network classification in conjunction with decision tree classifier, exploiting a global database of training sites interpreted from high-resolution Landsat TM imagery in association with ancillary data (Friedl et al., 2002). The algorithm of the latest collection of MODIS Land Cover products (MODIS Collection 5 Global Land Cover), includes an element to reduce year-to-year variability of classification results not associated with land cover change, particularly in classes that are ecologically proximate, due to poor spectral–temporal separability in MODIS data (Friedl et al., 2010). MODIS land cover products are available for every year in the period 2000-2012 as mosaics at a 500m spatial resolution or as spatially aggregated global datasets at 5' and 0.5° resolution.

Figure 11 compares MODIS land cover data for three years with the more detailed GlobeLand30 data. The figure shows that MODIS is fairly comparable between years, although differences in detail exist as is clear from the more limited occurrence of local variation in the 2006 image. The GlobeLand30 image shows much more agricultural activity in the central valley system of the Highlands. Both data sets are unable to discern the fragmented housing in the Highlands region.

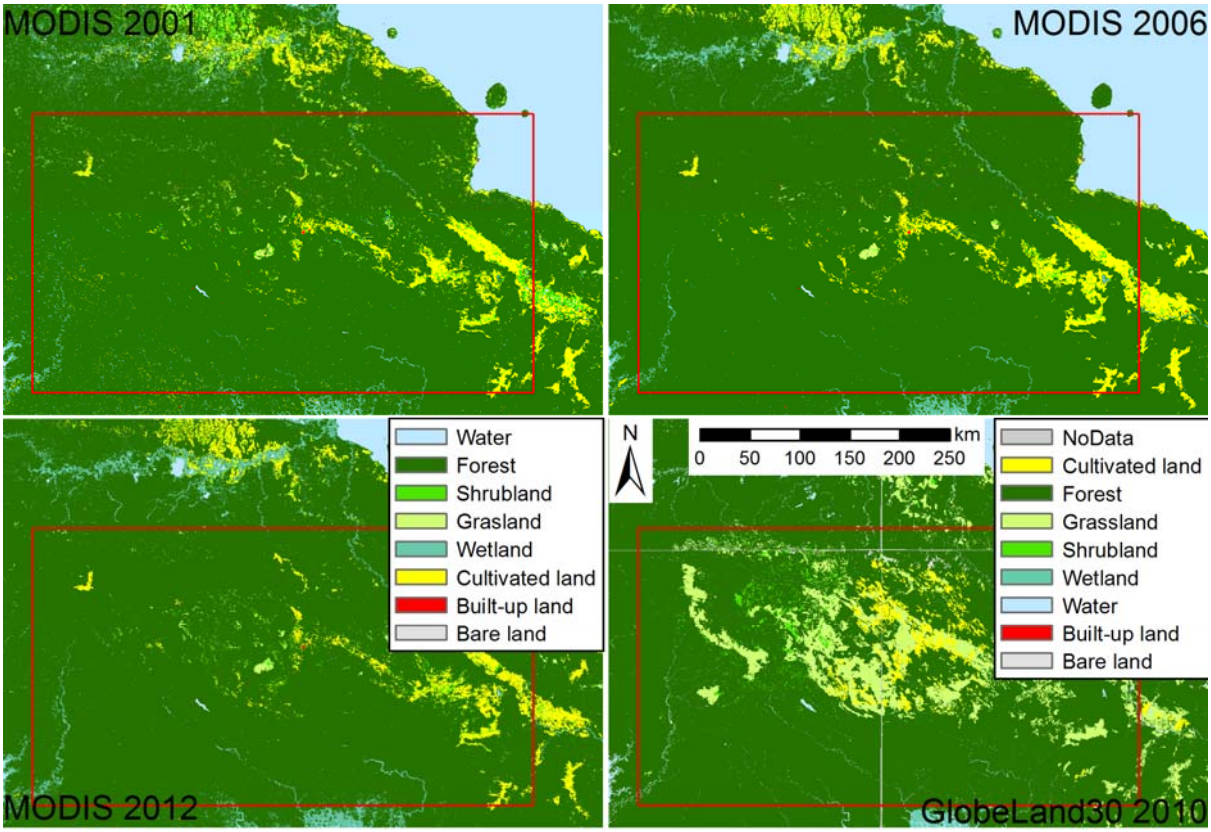

Figure 11 Land cover in the study area according to MODIS data for 2001, 2006 and 2012, compared with GlobaLand30 data. To allow comparison, the more detailed typology of the MODIS data has been aggregated to correspond to the GlobaLand30 description of land cover. Both images contain more spatial data then can be shown for this large area.

7.2. PNGRIS

An extremely valuable resource for spatially explicit information on the country's physical characteristics is provide by the Papua New Guinea resource Information System (PNGRIS). This extensive spatial database is the outcome of many land system studies in PNG by the Australian Commonwealth Scientific and Industrial Research Organisation (CSIRO) that have been carried out since 1953 (Vovola and Allen, 2001). It contains information on, amongst others, land form, rock type, soil type, altitude, relief, rainfall for each of the 4566 unique Resource Mapping Units (RMUs) distinguished for the country. An RMU is a relatively complex area of land characterised by a set of

natural resource attributes that are unique (Bellamy 1986). Information from this database can be used to control for differences in local physical characteristics that make some regions more productive and hence less poor than others. The spatial data sets in this database have been kindly made available by Bryant Allen from the Australian National University in Canberra. By way of example, Figure 1Figure 12 shows mean annual rainfall for most of Papua New Guinea from PNGRIS.

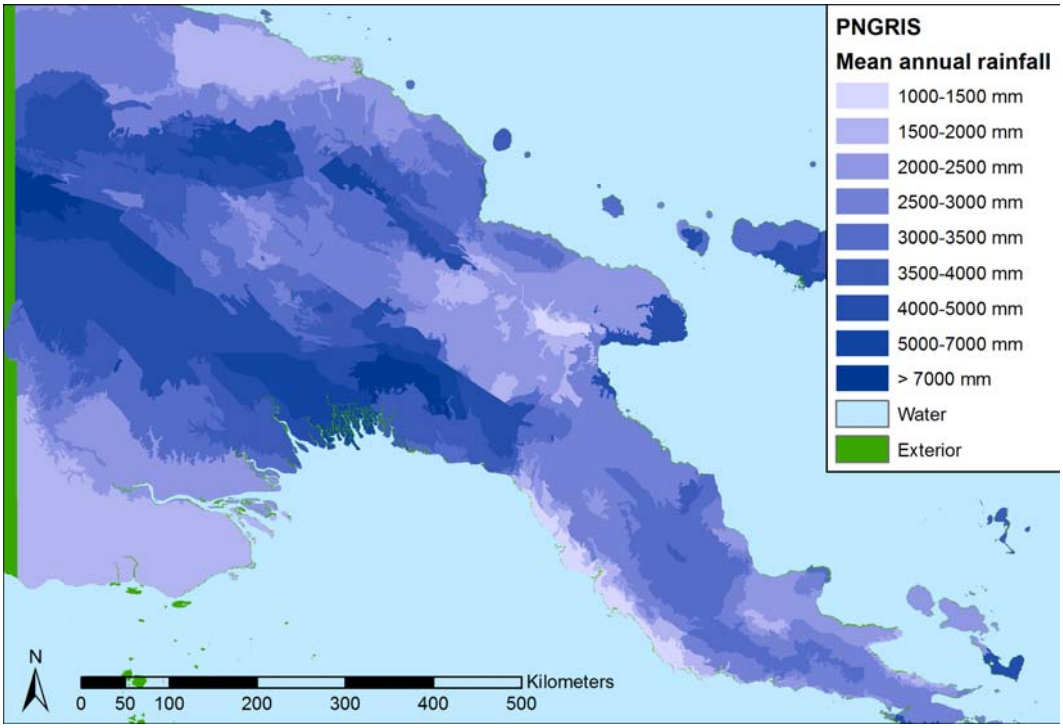

Figure 12 Mean annual rainfall (Source: PNGRIS).

8. Satellite imagery datasets

An inventory was made of open access sources of satellite imagery that can be helpful in mapping various aspects land use in Papua New Guinea. The data can be useful in addition to the land cover maps discussed in the preceding section to highlight specific features of the landscape. Section 9 discusses the (im)possibilities of a specific application of remote sensing data: mapping road quality.

8.1. Landsat

The most widely used repository of satellite imagery is the extensive Landsat archive that contains data from this family of satellites that have been used to map the earth since 1972. This data collection offers the best available source for time-series analysis of relatively long periods. Individual, geo-referenced images can be downloaded from the USGS earth explorer website (http://earthexplorer.usgs.gov/). Several images were downloaded for the Highlands region to explore their usefulness for the current study in mapping relevant spatial features (e.g. roads, see also Section 9). The data sets are selected for having a relatively low cloud coverage and approximating the most relevant years in our analsyis (2000 and 2010). As can be imagined for Papua New Guinea, fully cloud free images are virtually absent. Moreover, individual Landsat images measure about 200x200 km, so multiple images are required to cover the full Highlands region, making the collection and analysis of cloud scarce images for the whole area a time consuming task. This problem will even be larger for the more detailed commercial satellites that cover smaller areas.

A natural colour representation of three visible light bands (red, green and blue) of one of the included images is provided in Figure 13 and shows the level of detail (30m) available in these datasets. Specific operations exist to enhance the level of detail in the most recent Landsat images

(Landsat7 and 8, launched in 1999 and 2013 respectively) by including the 15m panchromatic band that represents the red, green, and blue portions of the electromagnetic spectrum in a grey scale. Unfortunately this is only possible for the Landsat8 images as it predecessor has suffered from problems with its sensor since 2003, producing striped images since then. So panchromatic sharpening is only possible from 1999-2003 and from 2013 onwards. In combination with the limited availability of cloud free images this implies that establishing full coverage of the study area with these sharpened images is most likely impossible.

Table 4 Specifications of the Landsat images included in the geo-database. For all images the full (Level 1) product is available as well ready to use LandsatLook Images with geographic reference.

Entity Id	Collection	Date
LT50990641997240ASA00	Landsat 5 TM	1997/08/28
L51ASA1011007000100_HDF	Landsat 5 TM	2011/01/07
LE70990642012354ASA00	Landsat 7 ETM (SLC-off 2003-present, causing stripes)	2012/12/19
LC80990642015226LGN00	Landsat 8	2015/08/14

Note: for additional metadata see, for example:
http://earthexplorer.usgs.gov/metadata/3119/LT50990641997240ASA00/ and
https://earthexplorer.usgs.gov/metadata/3373/LE70990642012354ASA00/

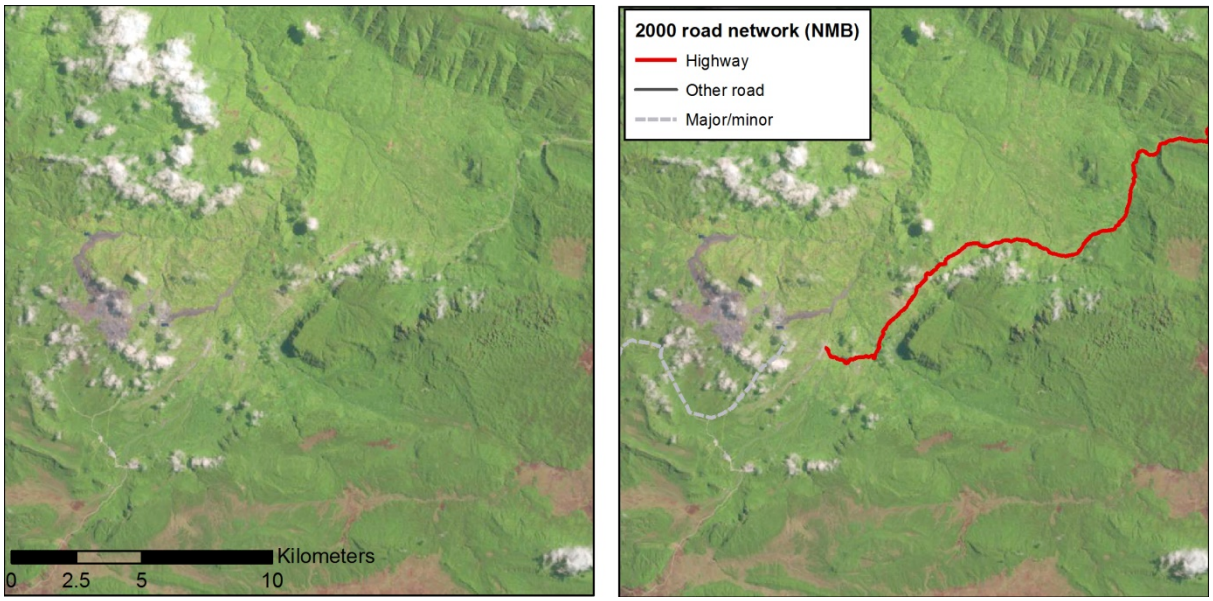

Figure 13 LandsatLook image (1997) compared to the roads classified by NMB around the Porgera gold mine. Comparison of the images reveals that resolution is too limited to properly recognise roads, let alone classify their type and quality. Section 9 discusses this topic in more detail.

8.2. Indian Remote Sensing Satellite (IRS)

The terms of reference for this project suggested the application of LISS-III data originating from the Indian Remote Sensing Satellite (IRS). Its latest mission (IRS-P6) is also known as ResourceSat-1, see https://earth.esa.int/web/guest/missions/3rd-party-missions/current-missions/irs-p6. After lots of trials and requests it turns out that only a limited set is available for some locations in Europe (through the EOLI-SA application distributed by the European Space Agency:-ESA). The Indian Government only distributes images for its own territory (http://bhuvan.nrsc.gov.in/data/download/index.php?c=s&s=L3&p=&g=). Not having access to this data source does not seem to hamper our analysis as the LISS-III satellite images have a resolution that is comparable to Landsat images (24m resolution as opposed to 30m for Landsat).

9. Mapping road quality

Part of our research aimed at investigating possibilities to supplement road quality data obtained from surveys with data obtained from satellite imagery. This section starts with an introductory treatment on remote sensing techniques applicable for automated road condition mapping. We describe the techniques commonly involved in such analyses, review some successful applications, and summarize data requirements and challenges related to current state of the art methods. A subsequent section concludes on the feasibility of automated road conditions mapping using satellite imagery. In Annex 1 we provide an extensive research paper that describes the methodology we developed to recognise roads based on Landsat images.

9.1. Remote sensing techniques for automated road condition mapping

Current practice in mapping road quality is by extensive field observations of experts that index local road quality based on a number of physical parameters, including cracking, ravelling, and rutting, (Herold et al., 2008). Other technologies, such as photographic and video logs of road quality coupled with GIS technology, are rapidly evolving. Such operations are, however, tedious and costly, and advances in remote sensing techniques could be of potential importance in supporting these infrastructure surveys (Brecher et al., 2004). While deriving physical and chemical properties on detailed level from hyperspectral imagery has been successful already in 1999 (Clark), mapping road quality from satellite imagery is still challenging depending on the goals and data available.

We distinguish different degrees in recognizable distress signals that categorize the severity of road deterioration. Each category links to a certain spatial resolution of the imagery data. On large spatial scales (low resolution), structural damages such as washouts, bridge collapses, and flooding damages have been successfully recognized by visual interpretation of satellite data and by Vector Change Analysis (VCA, see: Emery and Singh, 2013). VCA based algorithms subtract two images of an area and categorize the amount of change to recognize large structural changes.

Most research at lower spatial scales (high resolution) is directed at paved roads. Distress signals such as cracking can be recognized through image-processing techniques combined with a variety of image-recognition techniques (Cheng et al., 1999; Moussah and Hussain, 2001, Cheng et al., 2001; Chambon et al., 2009)[5]. Most of the images used in typical studies are however obtained at ground level, and emerging technologies focus on deploying Unmanned Airborne Vehicles (UAV, see, for example, Brecher et al., 2004; Zhang, 2010; Mei et al., 2014). Generally, aging and physical parameters of surface differentiation can be obtained from satellite imagery (Herold and Roberts, 2005; Mei et al, 2014), but the detail of imagery should be at least of sub-meter resolution with high detail in spectral bands. When imagery is somewhat less detailed (around 4m resolution) general categories of road quality and general indicators for condition have still been derived (Mohammadi, 2012), though other studies quote that efforts on linking spectral reflectance and physical characteristics at this resolution do not yield sensible results (Zhang and Elaksher, 2012). Some efforts in extracting road quality from imagery has focused on unpaved roads, but these studies rely on low altitude UAV (Zhang, 2008a; 2008b; 2010).

The general procedure in the efforts at different scale levels sketched above is similar. Aerial images are supplemented with highly detailed road quality surveys or *in-situ* measures, and spectral characteristics from the images are related to indicators or measures of road quality. After fitting a certain model that is able to link road condition and spectral characteristics, the model can be used to make out of sample predictions based on new image data.

[5] Recognition techniques include for example Support Vector Machines, Fuzzy Set Theory, Markov Methods and Neural Networks.

9.2. Feasibility

The advances in the past decades have made automated road quality mapping through combinations of image-processing techniques and image-recognition techniques a real option, but state of the art techniques require both highly detailed satellite imagery, highly detailed low altitude UAV imagery, or ground level imagery, supplemented with highly detailed road survey data or a richness of in-situ measures. As far as we know, none of these data sources is currently available for Papua New Guinea or should be obtained at very high costs. Sub-meter resolution imagery is available through Satellite Imaging Corporation or Landinfo (e.g. Quickbird, Worldview-1 & 3), Ikonos), with typical prices of around 15-60 US$ per square kilometer depending on the resolution[6]. Highly detailed road quality surveys or ground measurements, e.g. georeferenced road condition information for 10m road segments, is also not available. Taking ground measurements is a very tedious process and the study area is too large to obtain full coverage with a UAV. Moreover, performing such an incidental survey will only provide insight in the current status of the road network and not yield a time series of road quality surveys that can be used to link changes in household income over the past 20 years to changes in road quality. Unless the availability of high resolution imagery improves substantially, automated road quality mapping in the highlands region of PNG is currently is not feasible in the scope of this project.

Instead it seems more efficient to make use of road quality inventories that have been performed in preceding years. Below is a short overview of inventories that may be useful for the current project. Tracing information from these studies has high priority for the second phase of our project. A road asset management system (RAMS) was developed over 10 years ago by the Papua New Guinea Department of Works (DoW) and the Finnroad consultancy firm with funds and guidance provided by the Asian Development Bank (described in Jusi et al, 2003). Additionally an extensive Visual Road Condition Survey was performed by the Cardno company as part of the Australian Government-funded Papua New Guinea-Australia Transport Sector Support Program (TSSP) in cooperation with the Papua New Guinea Department of Works (www.pngtssp.com/index.php/projects/view/VRCS). This inventory combines video logs of road quality coupled with GIS technology and was expected to be finalized by 30 June 20157. Though this data cannot be directly linked to the 2009/10 household survey, it could make a valuable addition to our analysis.

The DoW plays a crucial role in maintaining roads and monitoring their quality. A recent quarterly road review report quotes percentages of roads falling in to certain quality categories and shows a map of missing links in the road network (www.businessmelanesia.com/wp-content/uploads/2014/DOW%20MAGAZINE%20ISSUE%2001.pdf). The report also includes maps and stories on recent plans on maintenance and new construction. The underlying data could be of crucial importance for our study, and might allow us to design powerful analysis methods. Unfortunately, the maps in the report are not readable.

Annex 1 documents some of our exploratory efforts in determining the usefulness of freely available Landsat imagery. Our conclusions are that the data is not suitable for automated procedures that extract vectorized road data or assign qualities variables to know roads. The lack of detail, coverage, and availability play an important role. However, Landsat imagery - if processed properly - can in some cases be interpreted by the human eye and be informative on ground level developments in the (unofficial) road network.

[6] For prices, see: http://www.landinfo.com/satellite-imagery-pricing.html

[7] This report seems to indicate that the TSSP data should be available and includes possible contacts: http://png.embassy.gov.au/files/pmsb/150625%20VRCS%20joint%20DoW%20AHC%20press%20release%20FINAL.pdf Additional information may be obtained from: http://www.cardno.com/en-au/projects/Pages/Visual-Road-Condition-Survey.aspx A possible contact for Cardno data is: http://people.bayt.com/barbara-lokes-22635573/

Part II Road improvement information

Based on information provided by DOW during our meeting in Mt. Hagen <<date>>we were able to map road improvements done between 2000 and 2014. This information was very useful but also a bit incomplete. For some stretches the year of completion was not mentioned, while some other stretches seem to have been left out (e.g. those financed by the National Government instead of the ADB). Moreover, the years that were listed as moment of improvement do not always correspond to the completion year listed in the Completion report of the Papua New Guinea -Road Maintenance and Upgrading (Sector) Project (that describes the status of the road improvements funded by ADB).

Therefore, we collected several reports and tried to make an overview of the upgraded road stretches (see table below).All improved stretches were digitized and added to our spatial data collection. They were used to calculate distances from sampled households to road stretches that were improved prior to the year of the survey2009. In subsequent statistical analysis we tried to establish differences in poverty level between households near upgraded roads as compared to those who did not receive upgrading. Unfortunately the number of observations was too limited to draw any statistically significant results.

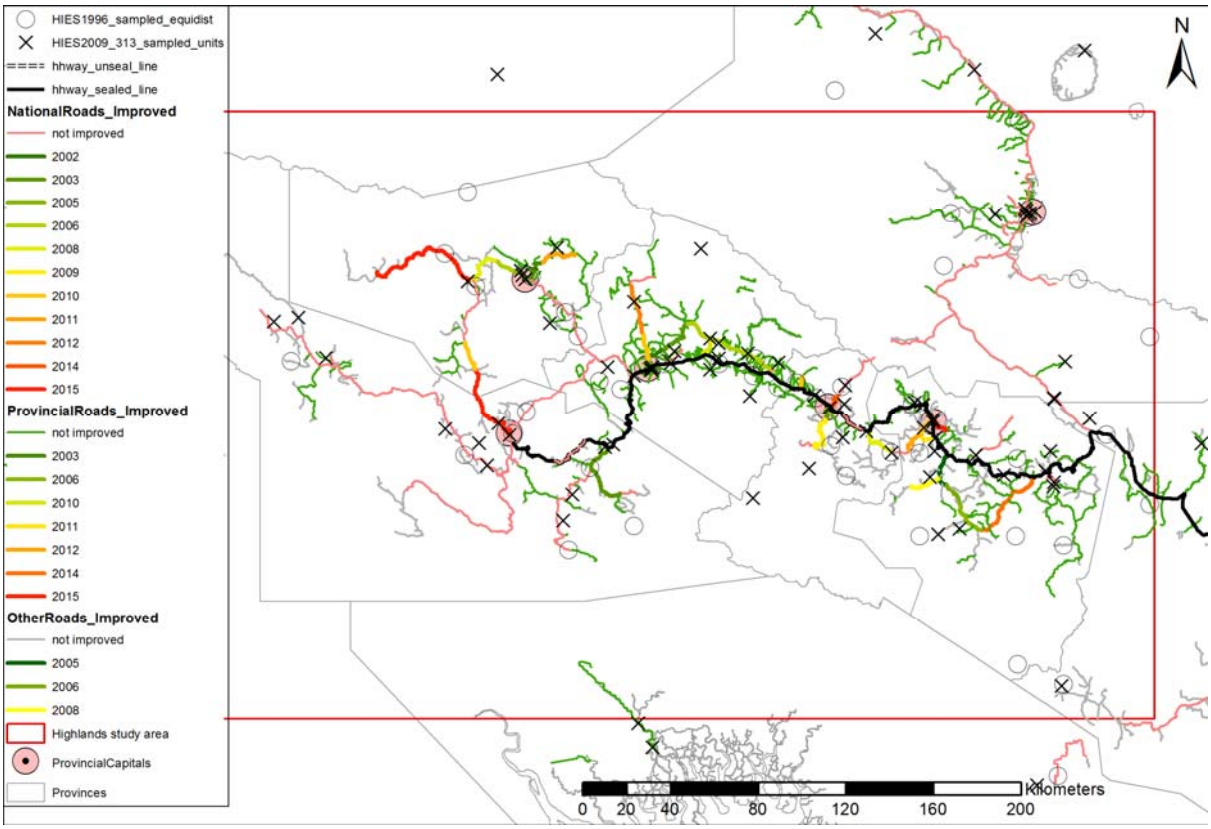

Figure 14 Improved road stretches in the Highlands region and sampled locations in the available household surveys.

Table 5 Overview improved roads as provided by DOW[1] during field visit, the corrected completion year was applied in constructing variables with distances to upgraded roads within the study period.

Nr.	Province	Road section	Completion year DOW[1]	Completion year ADB[2]	Completion year corrected[3]	Loan[4]	Road type	Remarks on selected completion year
1	Eastern Highlands	Korifegu-Nupuru	2005-2006	2010	2005	ADB-1709	Other road	Selected first year from two years provided by DOW, but completion report mentions 2010 (outside study period)
2	Eastern Highlands	Oliguti-Lufa	not listed	2008	2008	ADB-1709	Other road	Fully finalised in 2008 according to Completion report Appendix 7
3	Eastern Highlands	Nupuru-Okapa	2005-2006	>2007 (13% finalised) and <2013	2006	ADB-1709 -> Government	Other- & Provincial road	Selected second year from two years provided by DOW, but ADB Completion report (2013) states that only 13% was finalised in 2007 when project was handed over to Government who finalised it before 2013. So date could be later.
4	Eastern Highlands	Kamaliki-Bekuvia	not listed	>2009 (contracted) and <2013	2011	Supplementary loans	National- & Provincial road	Inferred from economic analysis (first year without cost in Appendix 11 = 2011), but could be later.
5	Eastern Highlands	Rypinka-Okapa	2014	>2013	2014	Supplementary loans	Provincial road	Completion report indicates that project was not completed by 2013
6	Chimbu	Kundiawa-Goro	not listed	2003	2003	ADB-1709	National road	
7	Chimbu	Kundiawa-Gewa	2014	not listed in report	2014	ADB-1709	National road	We travelled this stretch during field visit, not described in Completion report
8	Chimbu	Chuave-Movi	not listed	2008	2008	ADB-1709	National road	Inferred from economic analysis (first year without cost in Appendix 11 = 2008)
9	Chimbu	Koronige-Kerowagi	not listed	2010	2010	Supplementary loans	National road	Inferred from economic analysis (first year without cost in Appendix 11 = 2010)
10	Western Highlands	Ogelbeng-Ambra	2002	2002	2002	ADB-1709	National road	
11	Western Highlands	Ambra-Kotna	2003	2006	2003	ADB-1709	National road	Completion report states later data, but both fall within study period
12	Western Highlands	Kotna-Banz	2006	>=2006 (started)	2006	ADB-1709	National road	Completion report only mentions start of work not its completion
13	Western Highlands	Banz-Karameng	2006	2010	2006	Supplementary loans	National road	Completion report states later date (2010) that falls outside study period
14	Western Highlands	Karameng-Dona	2006	<2013	2006	Supplementary loans	National road	Listed as finalised in Completion report, no details listed in Appendix 7,
A	Southern Highlands	Kisenepoi-Ialibu	2003	2004 (90%)/2009	2003	ADB-1709	National road	Seems to have been finished 90% around 2004, last part delayed by new contracting (Appendix 12 ADB-Completion report)
B	Southern Highlands	Ialibu-Pangia	2003	>2009 (contracted) and <2013	2003	Supplementary loans	National road	Year completed seems unlikely for supplementary loan project. Completion report states other dates, for now we have assumed 2003.
C	Southern Highlands	Mendi-Kandep	not listed	not applicable	2015	HRRIIP	National road	Full completion expected in 2015 according to Resettlement Monitoring Report; 82% was completed by end 2014
D	Enga	Wabag-Warumanda	2005	2009	2005	ADB-1709	National road	Completion report states later date (2009) that almost falls outside study period
E	Enga	Warumanda-Andambi	2006		2006	ADB-1709	National road	This subsection seems to be part of the overall section from Wabag to Warumanda; it is not listed separately in Supplementary loans report
F	Enga	Warumanda-Laiagam	2008	<2013	2008	Supplementary loans	National road	Mentioned as Lame Bridge-Laiagam in most tables of Completion report Appendix 2 states it to be finished. Appendix suggest that work started after 2009
G	Enga	Laiagam-Porgera	not listed	not applicable	2015	HRRIIP	National road	Full completion expected in 2015 according to Resettlement Monitoring Report (nov.2014); 82% was completed by end 2014

Notes: [1]Information on upgraded road stretches and estimated year of completion provided by DOW (by Paul Nombri and Alphons Niggins) during field visit October 8, 2015.
[2]Completion year as stated in ADB Completion report (2013, Appendix 7).
[3]Year of completion is primarily based on specification of DOW, but for those cases where no year was specified we assumed a date based on other sources as specified in the motivation field. In some cases we relied on that information to correct the DOW information.
[4]Loan indication based on Supplementary loans project (ADB, 2006), sections C and D were found to be part of HRRIIP (Highlands Region Road Improvement Investment Program)

Table 6 Overview improved roads acoording to other sources[1]

Nr.	Province	Road section	Completion year ADB[2]	Completion year corrected	Loan	Road type	Remarks on selected completion year
15	Eastern Highlands	Goroka-Lahame		2015	Supplementary loans	Provincial road	Road was finished shortly before we travelled on it during field visit October 2015
16	Eastern Highlands	Goroka-Bekuvia	<2013	2010	Government funded	National road	Assumed finalised because listed as funded in Supplementary loan report of 2006
17	Enga	Wabag-Meriamanda	not applicable	2003	ADB-1709 -> Government	Provincial road	Completion report states stretch to have been finished with Government funding in 2003
18	Chimbu	Goro-Gumine	2003	2009	ADB-1709	National road	Completion report mentions that contracts were awarded in 2007 and road was completed by end of Program (2013). We have taken completion year from Appendix 12
19	Enga	Kaipau-Kandep	>2007 (contracted) and <2013	2010	ADB-1709	National road	Liagam-Kandep was shortened to Kaipau-Kandep
20	Western Highlands	Kindeng-Kondopina	2010	2010	Supplementary loans	Provincial road	
21	Western Highlands	Mt.Hagen-Baiyer road	2010	2010	ADB-1709 -> Government	National road	Taken over by Government in 2007. Full stretch finalised before 2013. Mt.Hagen-Notre Dame School subproject finished in 2010. Completion year remainder estimated at 2012
22	Enga	Meriamanda-Kompiam	>2010 (Mt. Hagen-N.D. School) and <2013	2012	ADB-1709 -> Government	Provincial road	Also completed with government funding. Assumed completion in 2012, but could also be other year.
	Southern Highlands	Mendi-Nipa	<2013	not finished	ADB-1709 -> HRRIIP		Initially part of ADB-1709, but terminated with only small section complete. Transferred to Government and included in HRRIIP

Notes: [1]This table lists road sections that were not mentioned by DOW but that were also upgraded as part of ADB's supplementary loans scheme.
[2]Completion year as stated in ADB Completion report (2013, Appendix 7).

During our field visit we obtained several spatial data sets describing the road network from the National Statistics Office (NSO) and the Department of Works (DOW). In addition, Bryant Allen (Australian National University) kindly provided a version of the original Road Asset Management System (RAMS) data set. Several datasets include reference to road type and condition, but contain limited information on their origin or the years they were created. After careful deliberation and discussion with ADB-officials we decided to use the datasets described below in our analysis.

RAMS data set on national and provincial roads in 2000

The original RAMS data sets describe both national roads (ca 6,437 km in total) and provincial roads (ca 7,404 km). The national roads dataset contained an impressive array of road quality variables (dated 1999-2001) including an overall characterization of road condition (in good, fair or poor). The dataset referring to provincial roads contained a similar set of road quality variables (dated 2001) but lacked the overall characterization of road condition. For the latter dataset we imputed a general road quality assessment based on the relation between several detailed quality aspects (e.g. severe damage, corrugation, IRI) and the overall assessment in national dataset.

NSO data set

From NSO we obtained a dataset with an extensive description of the road network. We have little information on its origin, but the names of the file indicate that it dates to 2009. The map includes the PNG road network in great detail (including tracks) as well as an assessment of road conditions. This possibly corresponds to data from the Road Asset Management System (RAMS). The reliability of RAMS data after its initial set up (around 2000) was questioned by Alphonse Niggins of the Department of Works (DoW) during our field visit. It seems no complete inspection of the road network was carried out before the start of the Visual Road Condition Survey in 2015. So the 2009 NSO data seems to follow from a more qualitative assessment of road quality. As this is our only source of road quality data that can be linked to the 2009/10 HIES data we decided to use this dataset in subquent analysis.

Building a complete road data set for 2000 and 2009

We believe that the higher density of roads depicted in the 2009 road map is to a large extent a result of an improvement of the information contained in the map, rather than the construction of new roads. This is confirmed by a report of the World Food Programme and Logistics Cluster (WFP, 2011) that states that between 2000 and 2010 the focus of road works was on maintenance and upgrading, but no new roads were built. To create a complete representation of the 2000 network we therefore matched the road attributes of the 2000 data sources to the NSO road data set.

The road quality attributes for 2000 were linked to the NSO road dataset based on the road section ID where available. For the remaining road stretches the road attributes where matched on spatial proximity (based on the length of intersection of non-matched national NSO roads stretches with a 200m buffer around national roads and a one-to-many spatial join of the non-matched provincial NSO roads with the attributes of the provincial roads within a 200m distance). The spatial joins typically provide matches with multiple road stretches, in which cases we selected the longest overlaps for the match of attributes. This process leads to the matching of all info on the national roads from the RAMS data and most of the info from the provincial roads. Only when roads are classified differently (e.g. as National roads in the NSO data and provincial roads in the RAMS data) this does not seem to work. For these stretches and the approximately 12,000 (!) km of other roads that exist in the NSO data the conditions of 2010 are used as proxy for the 2000 roads, thus assuming road quality did not change over time.

The table below summarises the classifications of surface type and road condition in both years, while the two subsequent figures show the road condition

Table 7 Surface type and road conditions as described in the two road network data sets

Surface	**Condition**	**2009 (NSO data)**		**2000 (RAMS matched to NSO)**	
		Total length (km)	**%**	**Total length (km)**	**%**
Dirt	Poor	4332	17.0%	1230	10.5%
Gravel	Poor	5726	22.4%	4232	36.3%
Sealed	Poor	371	1.5%	314	2.7%
Dirt	Fair	3660	14.3%	63	0.5%
Gravel	Fair	7300	28.6%	1649	14.1%
Sealed	Fair	1067	4.2%	914	7.8%
Dirt	Good	166	0.7%	223	1.9%
Gravel	Good	1096	4.3%	2137	18.3%
Sealed	Good	1799	7.0%	911	7.8%
Total		*25,517*	*100%*	*11,672*	*100%*

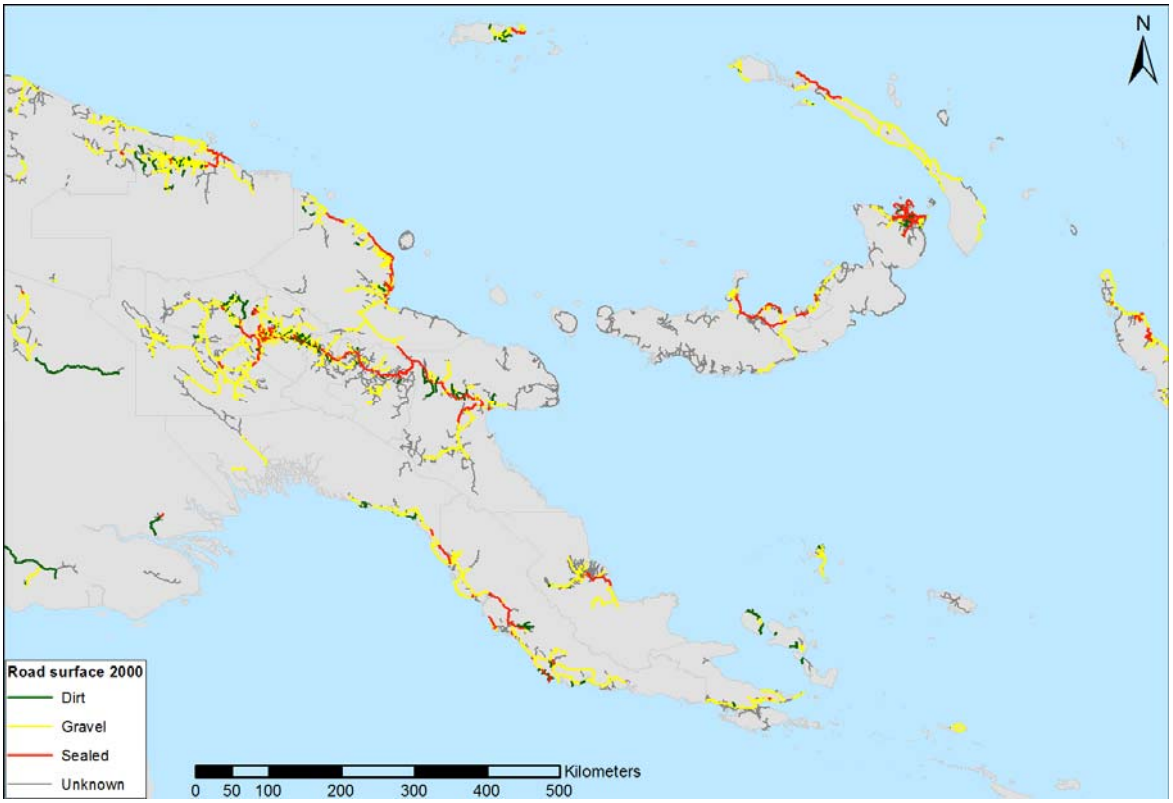

Figure 15 Road surface type 2000 based on the combination of attribute information from the RAMS data and spatial data describing road location from NSO. Road surface type is unknown for road stretches that were not included in RAMS data. These are typically minor dirt roads not managed by state or province.

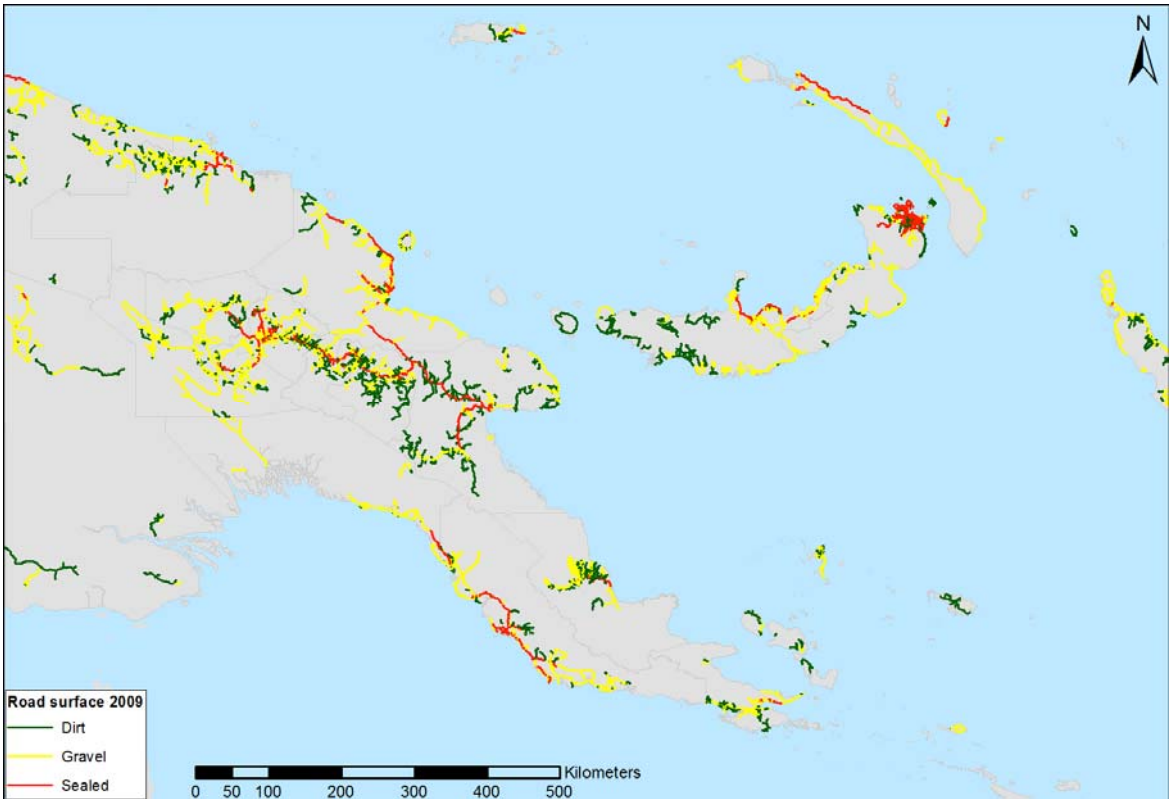

Figure 16 NSO road network with indication of road surface type.

Part III Network analyses to calculate road-based distances

GIS-based network analyses were applied to create explanatory variables with travel distance from the household survey units to the nearest town. As the NSO road data contains the most extensive and recent assessment of road quality we apply that data set to build set up our network analysis. This requires a number of data processing and analysis steps that are described below. This description is fairy technical, but included here to provide insight in applied methodology and choices that were made. Essentially we followed the following steps in ArcGIS 10.4:

1. Conversion of matched NSO data (NSO_roads_2010_2000_matched_EQdist.shp) to a projected coordinate system to be able to calculate distances in meters (project to Asia_South_Equidistant_Conic);
2. Repair the network. To create a meaningful network all lines representing roads should be properly connected. Unfortunately many, many small gaps exist in the network (typically around 10 or so meters between road stretches) that have to be corrected. Doing this manually for mainland PNG took too much time as thousands of road stretches had to be corrected. Therefore a more automated approach was applied: selecting many road segments at the same time and applying the advanced editing option 'planarise lines'. This adds nodes at intersections and moves nodes within a predefined range (for which 10 metres was selected).
3. Small corrections on data set. Two tiny stretches were deleted (1.3 km in total) without attribute information for 2010. In addition the spelling was harmonized in surface type (dirt/Dirt -> Dirt; Gravel/GRAVEL-> Gravel) and Condition (fair/Fair->Fair) in the attributes related to 2010 and 2000.
4. Add an attribute related to travel cost to the road network (called impedance in ArcGIS, for example expressed in road length or travel time). In this case we added meters (using calculate geometry). As we also specify road type and condition per road stretch we can later summarise the length travelled over different types of roads and conditions.
5. Specify a basic network for network analysis without turns, using any vertex connectivity, no elevation or travel mode and specify the different length attributes as cost attributes.
6. Calculate distance from each census unit to the nearest town using the Closest Facility option. Some specifications: load census units as locations for Incidents and 15 km as search tolerance (i.e. census unit will only be linked to the nearest road when that is within 15km); load selection of towns with more than 1000 inhabitants according to the census 2011. In this case specifying a maximum distance of 5km to reach this town from a road.

A simple example of the network-based distances this process generates is included in Figure 17. The figure shows the shortest routes (thick green line) over the road network (thin line) linking census units and towns. The results can be stored as a table and we specified the analysis in such a way that the we record not only the total length of the route in metres, but also the total length travelled on different types of roads distinguishing unique combinations of surface type (sealed, gravel, dirt) and road condition (good, fair, poor) in both 2000 and 2009 (so 3 x 3 x2 = 18 attributes[8]). Note that these values only include distances over the network. The overland distances are added to that based on simpler GIS-based calculations that look for the nearest road from each sampled census unit.

[8] In fact we created more attributes as we also distinguished between the known road conditions in 2000 and the assumed ones (where the missing quality information was filled in by taking the 2009 conditions). These different distance calculations are used as alternative specifications in final paper describing the statistical analysis.

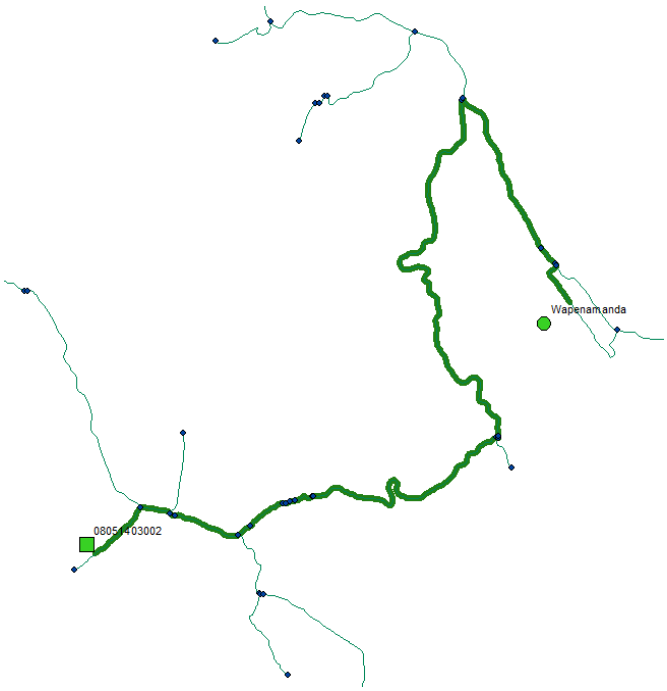

Figure 17 Example of network-based distance calculation between a sampled household (indicated by number) and nearest town (Wapenamanda).

After this initial analysis a few unrealistic long routes and not-connected census units were found that were related to obvious data gaps (e.g. roads missing in larger towns that were not part of the RAMS data). These missing links were added to the network, as were the major roads for Port Moresby to allow census units in the surrounding rural areas to link to this major town. The final results now present a convincing representation of the routes villagers would take to go to town. The current analysis may still overestimate travel distance due to network inconsistencies (when connections miss that should have been present), but we are confident that the most important connections are now included. In fact, an underestimation of the difficulty of travel may also be present as some road stretches may not be passable at all times due to seasonal conditions or incidental road failure.

Some key statistics of the curent analysis: 47 out of 50 towns are linked to the road network; 297 of the 313 sampled census units from the 2009/2010 survey are linked to a road. All remaining places are either located along the coast are deep in the interior and indeed seem too far from a road to be connected. Some additional census units are without a link to a town because the roads they are located on do not reach a town (e.g. because of water bodies). In some cases ferry services may exist that provide a connection. But as we do not know have information about that, we consider these to be missing (or at least much poorer) connections. This may of course have some impact on the results for the islands, although roads may be less important here as water traffic may be more prevalent. Figure 18 below shows the non-connected census units and towns in a lighter colure with a question mark and the shortest routes between connected census units and towns are purple lines.

It is also possible to create a full origin-destination matrix linking each census unit to all town reachable over the road network starting the census unit. This results in a large Origin-Destination matrix of 3309 cells (not all origins link to all towns). Figure 18 depicts all established connections as straight lines and clearly shows the disconnected larger regions of the country. These distance calculations analysis can be used to add second nearest destinations to the analysis, or, for example,

include the distances to specific (larger) towns such as provincial capitals. In the final, statistical analysis of this project they were not used.

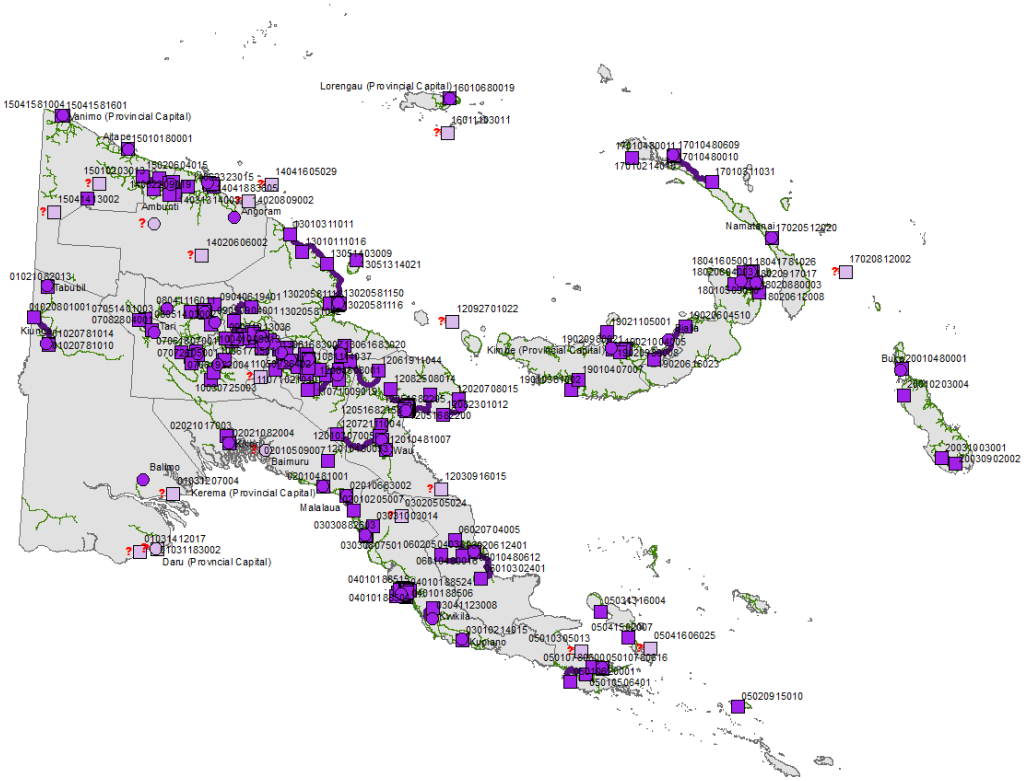

Figure 18 Network analysis for Papua New Guinea showing sampled census units as squares, towns as circles and the shortest routes between them as purple lines. Question marks denote census units or towns that could not be located to a road stretch within the specified search distances.

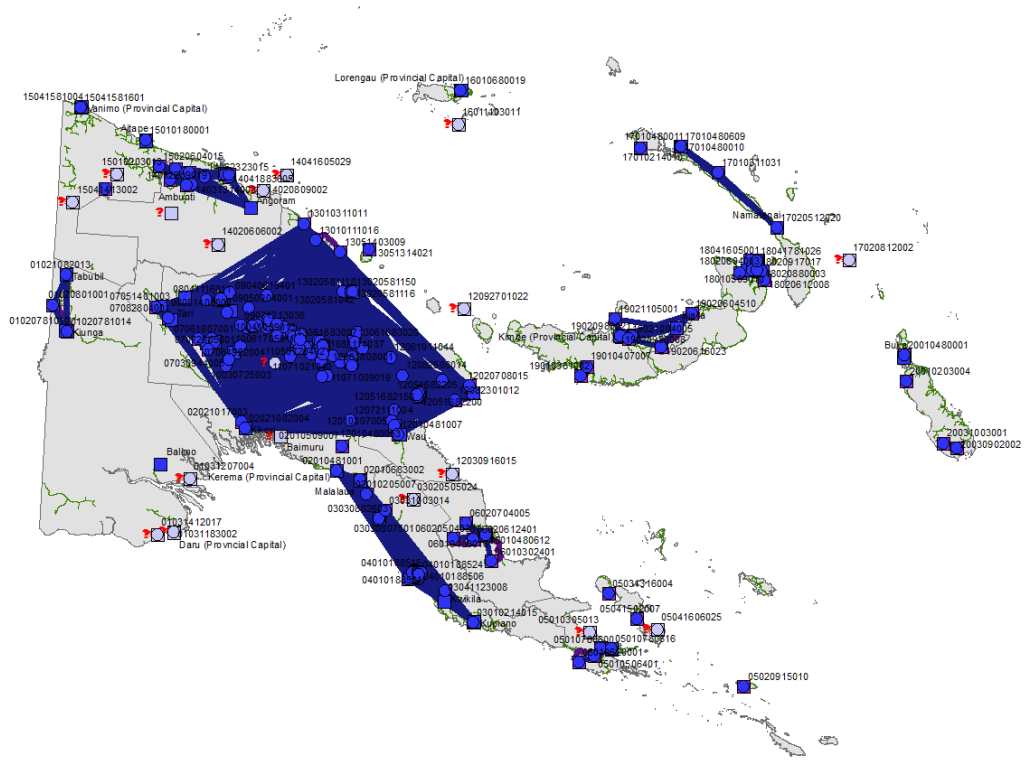

Figure 19 Scematic represenattion of all possible connections between sampled census units and towns.

References

ADB (2006) Proposed Supplementary Loans. Papua New Guinea: Road Maintenance and Upgrading (Sector) Project. Asian Development Bank.

ADB (2013) Completion report. Papua New Guinea: Road Maintenance and Upgrading (Sector) Project. Asian Development Bank.

ADB (2015) Resettlement Monitoring Report Papua New Guinea: Highlands Region Road Improvement Investment Program – Mendi to Kandep Road. Asian Development Bank.

Balk, D.L., Deichmann, U., Yetman, G., Pozzi, F., Hay, S.I., Nelson, A. (2006) Determining global population distribution: Methods, applications and data. Advances in Parasitology 62: 119–156.

Bellamy, J.A. (1986) Papua New Guinea Inventory of Natural Resources Population Distribution and Land Use: Handbook. Natural Resources Series No. 6. Melbourne, Australia, Division of Water and Land Resources, Commonwealth Scientific and Industrial Research Organisation.

Bhaduri, B., Bright, E., Coleman, P., Urban, M.L. (2007) LandScan USA: A high resolution geospatial and temporal modeling approach for population distribution and dynamics. GeoJournal 69: 103-117.

Brecher, A., Noronha, V., Herold, M. (2004) UAV2003: A roadmap for deploying unmanned aerial vehicle (UAVs) in transportation, Findings of Specialist Workshop in Santa Barbara, CA, December 2003. Downloaded from: http://www.ncgia.ucsb.edu/ncrst/meetings/20031202SBA-UAV2003/Findings/UAV2003-Findings-Final.pdf (accessed: September 2015).

Chambon, S., Subirats, P., Dumoulin, J. (2009) Introduction of a Wavelet Transform Based on 2D Matched Filter in a Markov Random Field for Fine Structure Extraction: Application on Road Crack Detection. IS&T/SPIE Electronic Imaging, Image Processing: Machine Vision Applications II, San Jose, USA.

Chen, X., Nordhaus, W. (2015) A Test of the New VIIRS Lights Data Set: Population and Economic Output in Africa. Remote Sensing 7: 4937-4947.

Cheng, H.D., Chen, J.R. , Glazier, C.,Hu, Y. G. (1991) Novel Approach to Pavement Cracking Detection Based on Fuzzy Set Theory. Journal of Computing in Civil Engineering 13 (4): 270-280.

Cheng, H.D., Wang, J.L., Hu, Y.G., Glazier, C., Shi, H.J., Chen, X.W. (2001) Novel Approach to Pavement Cracking Detection Based on Neural Networks, Transp. Res. Record 1764: 119-127.

Clark, R.N. (1999) Spectroscopy of rocks and minerals and principles of spectroscopy. Chapter 1 in: Rencz, A.N . (Ed.), Manual of Remote Sensing. John Wiley and Sons, New York: 3-58.

Dobson, J.E., Bright, E.A., Coleman, P.R., Durfee, R.C., Worley B.A. (2000) A global population database for estimating population at risk. Photogrammetric Engineering and Remote Sensing 66 (7): 849-857.

Doll, C.N.H. Muller, J., Morley, J. (2006) Mapping regional economic activity from night-time light satellite imagery. Ecological Economics 57 (1): 75-92.

Doll, C.N.H., Muller, J.P., Elvidge C.D. (2000) Night time imagery as a tool for global mapping of socioeconomic parameters and greenhouse gas emissions. AMBIO: A Journal of the Human Environment 29 (3): 157-162.

Doll, C.N.H., Pachauri, S. (2010) Estimating rural populations without access to electricity in developing countries through night-time light satellite imagery. Energy Policy 38 (10): 5661-5670.

Elvidge, C.D., Baugh, K.E., Kihn, E.A., Kroehl, H.W., Davis, E.R. (1997) Mapping city lights with the night time data from the DMSP operational linescan system. Photogrammetric engineering and remote sensing 63 (6): 727-734.

Elvidge, C.D., Imhoff, M.L., Baugh, K.E., Hobson, V.-R., Nelson, I., Safran, J., Dietz, J.B., Tuttle, B.T., (2001) Night-time lights of the world: 1994-1995. ISPRS Journal of Photogrammetry & Remote Sensing 56: 81–99.

Elvidge, C.D., Sutton, P.C., Ghosh, T., Tuttle, B.T. Baugh, K.E., Bhaduri, B., Bright, E. (2009) A global poverty map derived from satellite data. Computers & Geosciences 35 (8): 1652-1660.

Emery, W., Sing, C. (2013) Large-Area Road-Surface Quality and Land-Cover Classification Using Very-High Spatial Resolution Aerial and Satellite Data. Final Report for US DOT DTPH56-06-BAA-0002.

Friedl, M.A., McIver, D.K., Hodges, J.C.F., Zhang, X.Y., Muchoney, D., Strahler, A.H.,Woodcock, C.E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., Schaaf, C. (2002), Global land cover mapping from MODIS: algorithms and early results. Remote Sensing of Environment 83: 287–302.

Friedl, M.A, Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., Huang, X. (2010), MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. Remote Sensing of Environment 114 (1): 168–182

Fischer, M.M. and A. Getis (eds) 2010 Handbook of Applied Spatial Analysis, Software Tools, Methods and Applications, Springer.

Gaughan, A.E., Stevens, F.R., Linard, C., Jia, P., Tatem, A.J. (2013) High Resolution Population Distribution Maps for Southeast Asia in 2010 and 2015. PLoS ONE 8(2): e55882.

Ghosh, T. Sutton, P. Powell, R. Anderson, S., Elvidge, CD. (2009) Estimation of Mexico's informal economy using DMSP night time lights data. Remote sensing 1: 418-444.

Ghosh, T., Powell, R., Elvidge, C. D., Baugh, K. E., Sutton, P. C., Anderson, S. (2010) Shedding light on the global distribution of economic activity. The Open Geography Journal (3): 148-161.

Gibson, J. and Rozelle, S. (2003) Poverty and Access to Roads in Papua New Guinea. Economic Development and Cultural Change 52 (1): 159-185.

Gibson, J., Datt, G., Allen, B., Hwang, V., Bourke, R.M., Parajuli, D. (2005). Mapping Poverty in Rural Papua New Guinea, Pacific Economic Bulletin 20(1): 27-43.

Herold, M., Roberts, A. (2005) Spectral characteristics of asphalt road aging and deterioration: implications for remote-sensing applications. Applied Optics 44 (20): 4327-4334.

Herold, M., Roberts, D., Noronha, V., Smadi, O. (2008) Imaging spectrometry and asphalt road surveys. Transportation Research Part C 16: 153–166.

Imhoff, M.L., Lawrence, W.T., Stutzer, D.C., Elvidge, C.D. (1997) A technique for using composite DMSP/OLS “City Lights” satellite data to accurately map urban areas. Remote Sensing of Environment, 61 (3): 361-370.

Jacoby, H. G. (2000), Access to Markets and the Benefits of Rural Roads. The Economic Journal, 110: 713–737.

Jusi, P., Mumu,R., Jarvenpaa, S., Nease, B., Sangrador, E. (2003) Road Asset Management System Implementation in Pacific Region: Papua New Guinea. Transportation Research Record 1819: 323-332.

Levin, N., Duke, Y. (2012) High spatial resolution night-time light images for demographic and socio-economic studies. Remote sensing of Environment 119: 1-10.

Mei, A., Salvatori, R., Fiore, N., Allegrini, A., D’Andrea, A. (2014) Integration of Field and Laboratory Spectral Data with Multi-Resolution Remote Sensed Imagery for Asphalt Surface Differentiation. Remote Sensing 2014 (6): 2765-2781.

Meyer, D. (ed.) (2011) ASTER Global Digital Elevation Model Version 2 – summary of validation results. NASA Land Processes Distributed Active Archive Center and the Joint Japan-US ASTER Science Team. Downloaded from: http://asterweb.jpl.nasa.gov/gdem.asp (accessed: September 2015).

Mills, S., Weiss, S., Liang, C. (2013) VIIRS day/night band (DNB) stray light characterization and correction. Proc. SPIE 8866, Earth Observing Systems XVIII, 88661P.

Mohammadi, M. (2012) Road Classification and Condition Determination using Hyperspectral Imagery. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XXXIX-B7, 2012 XXII ISPRS Congress, 25 August – 01 September 2012, Melbourne, Australia.

Nordhaus, W., Azam, Q., Corderi, D., Hood, K., Victor, N.M., Mohamed, M., Miltner, A., Weiss, J. (2006) The G-econ Database on gridded output: Methods and Data. Technical report, Yale University.

Stevens, F.R., Gaughan, A.E., Linard, C., Tatem, A.J. (2015) Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. PLoS ONE 10(2): e0107042.

Sutton, P.C., Elvidge, C.D., Ghosh, T. (2007) Estimation of gross domestic product at sub-national scales using nighttime satellite imagery. International Journal of Ecological Economics & Statistics 8 (SO7): 5-21.

Vovola, P., Allen, B. (2001) Computer managed databases relevant to PNG agriculture. In: Bourke, R.M., Allen, M. G., Salisbury, J.G. (eds.) Food Security for Papua New Guinea: Proceedings of the Papua New Guinea Food and Nutrition 2000 Conference, Lae PNG, Australian Centre for International Agricultural Research, Canberra: 467-475.

WFP (2011). Papua New Guinea Emergency Preparedness: Operational Logistics Contingency Plan Part 2 - Existing Response Capacity and Overview of Logistics Situation. World Food Programme (WFP) Logistics Cluster, Rome.

Welch, R. (1980) Monitoring urban population and energy utilization patterns from satellite data. Remote Sensing of Environment 9: 1-9.

Zhang, C. (2008a) Development of a UAV-based Remote Sensing System for Unpaved Road Condition Assessment. Proc. of ASPRS Annual Conference. April 28 – May 2, Portland, OR.

Zhang, C. (2008b) Development of a UAV-based Remote Sensing System for Unpaved Road Condition Assessment. International Achieve of Photogrammetry and Remote Sensing. July 3-13, Beijing, China.

Zhang, C. (2010) Assessment of Rural Road Condition using UAV-based Remote Sensing. Proc. of Remote Sensing Technologies for Transportation Applications. Washington DC.

Zhang, C., Elaksher, A. (2012) Development of an Unmanned Aerial Vehicle-Based Imaging System for 3D Measurement of Rural Road Surface Distresses. Computer-Aided Civil and Infrastructure Engineering. Computer-Aided Civil and Infrastructure Engineering 27 (2): 118–129.

Annex 1 Automated recognition of unpaved roads in rural Papua New Guinea using Landsat 8: Spectral content analysis

Research paper describing the methodology we developed to automatically recognise unpaved roads in rural Papua New Guinea. This paper may be further expanded into a journal publication.

Automated recognition of unpaved roads in rural Papua New Guinea using Landsat 8: Spectral content analysis

Bo Pieter Johannes Andrée[1,*]

[1] *Department of Spatial Economics/SPINlab, VU Amsterdam, Netherlands*
[*] *Corresponding Author: email b.p.j.andree@vu.nl*

February 23, 2016

Abstract

I explore wavelength properties of different land-cover classes to understand how Landsat 8 imagery can be used for automated road recognition in developing countries with tropical conditions. Distinguishing roads from other barren surfaces constitutes the main challenge, and I show that differences in spectral content are relatively small between these surfaces while stronger w.r.t. forest cover. Radiance curves are similar among different barren surfaces, and detail in spectral bands is insufficient to distinguish different barren land types based on black body temperatures. False colour composites however, show that in the near infrared and short infrared wavelengths, roads can be distinguished from other types of barren land. In general roads have higher reflectance throughout the whole visible spectrum, but there is also larger variation in DN. Furthermore, the deep blue waveband shows promising properties for unpaved road detection. However, tropical regions are plagued by atmospheric distortion and its usefulness for road detection is likely to be limited to higher altitutes. In the investigated study area, this does not pose a threat as road occurrence intersects with high altitude terrain. Finally, I propose an algorithm based on high-pass filters that is able to enhance visualization of roads and allows to extract information that is not visible in pan-sharpened imagery. The final recommendation for development of road recognition algorithms that work with Landsat 8, is to rely on bands 1-8, focus on high altitude terrain, and feed on the output of the high-pass algorithm. Training samples should not include air strips or washouts due to differential properties in near- and shortwave infrared.

Keywords: Remote Sensing, Automated Road Mapping, Road Recognition, Papua New Guinea, Spectral Reflectance Curves, Image Processing, Landsat 8.

1 Introduction

In past years, there have been numerous efforts to maintain and improve the road network in the Highlands region in Papua New Guinea (PNG). The Asian Development Bank (ADB) has contracted the VU to explore potential economic impacts of road development programs in PNG. A recurring problem throughout the project has been the scarcity of accurate data on road networks. Remote sensing techniques have the potential to complement current incomplete datasets through automated road detection. It is therefore useful to the ADB, and future researchers, to know to what extend automated road detection is possible using only freely available satellite imagery.

The main purpose of this research is to obtain insights that will aid researchers develop automated road detection procedures for areas similar to PNG. In this short paper I will explore spectral content of Landsat 8 imagery. The focus is to contrast the spectral characteristics of two types of barren land, with those of identified unpaved road surfaces. Since both barren land and unpaved roads have similar soil properties and similar radiance properties in the visible wavelengths, exploiting minute differences between these types of land cover can be expected to constitute the main challenge in automated road detection in PNG. Additionally I investigate forest land cover to see how differences between barren surfaces and unpaved roads relate to differences between land-cover classes that can be more easily distinguished. Next to analysing spectral content, I propose an algorithm that makes use of high-pass filters to further enhance pan sharpened images. The output of the algorithm could support road recognition.

The remaining part of this paper is as follows. Section 2 briefly describes the study area, section 3 provides several color displays that provide initial insights regarding the spectral content of different land-cover types. In section 4 I explore radiance statistics of all Landsat 8 bands. In section 5 I investigate ancilarry data. I first combine a pan-sharpened true color composite with a digital elevation model (DEM) obtained from the ASTER satellite, and overlay this map with road vector data and population data to better understand how the road network is positioned in the geo-economic landscape. I then propose the high-pass filter based algorithm and explore its potential to extract additional information usable in a four channel pan-sharpened image. Finally, section 6 briefly summarizes key results regarding the usability of Landsat 8 for recognition of unpaved roads in tropical areas.

2 Describing the study area

The area chosen for this particular study is an excerpt of the eastern part of the New Guinea Highlands (the Highlands) region in PNG. The Highlands comprises a chain of

mountain ranges and river valleys that run east-west over the length of the entire island of New Guinea. PNG is known as a late-developing country in which vast parts of the land remained isolated from the world until well after 1950. Because of it's mountainous terrain, the country has had difficulty in connecting and maintaining it's fragmented transportation network, particularly in the Highlands region. The eastern part of the Highlands is at the same time relatively densely populated by poor agricultural communities. The non-uniform distribution of road access, and the weak local economy, provide a unique opportunity to study local effects of connectedness on poverty levels. However, accurate time-series data on the road network is key in such a study, and programms for continuous maintenance of such datasets have been failing due to structural instability of governmental institutions. Therefore the study area provides a unique opportunity for policy interest and challenges in remote sensing techniques to meet in one study.

Figure 1 positions the imagery of the study area within the central part of PNG. Due to tropical conditions, cloud-free data is limited for PNG, but the specific image chosen captures the most important differences in landscape elements in an area that is interesting for policymakers. It includes:

(i) Several mining sites and air strips that are unpaved and of artificial nature;

(ii) Various types of unpaved roads;

(iii) Rivers of different sizes;

(iv) Various types of natural unpaved terrain;

(v) A wide range of altitudes, rugged terrain and a flat lowlands.

The presence of these different factors allows the exploration of differences in spectral content of various unpaved terrain types along an altitude gradient. Furthermore, the particular image set allows a comparison between rivers and roads, which is interesting for road extraction algorithms that rely on linear feature detection such as in Gecen and Sarp (2008).

A more detailed visual inspection of the study area is reserved for the next section, in which I will present several colour displays. From figure 1 it can however already be seen that the southern part of the area includes significantly more clouds, whereas the northern part contains misty inter-mountain river valleys. The middle part of the image is mist- and cloud free. To limit the scope of this research, the remainder part of this paper will mostly focus on this central part of the depicted image set where atmospheric distortion is at it's lowest.

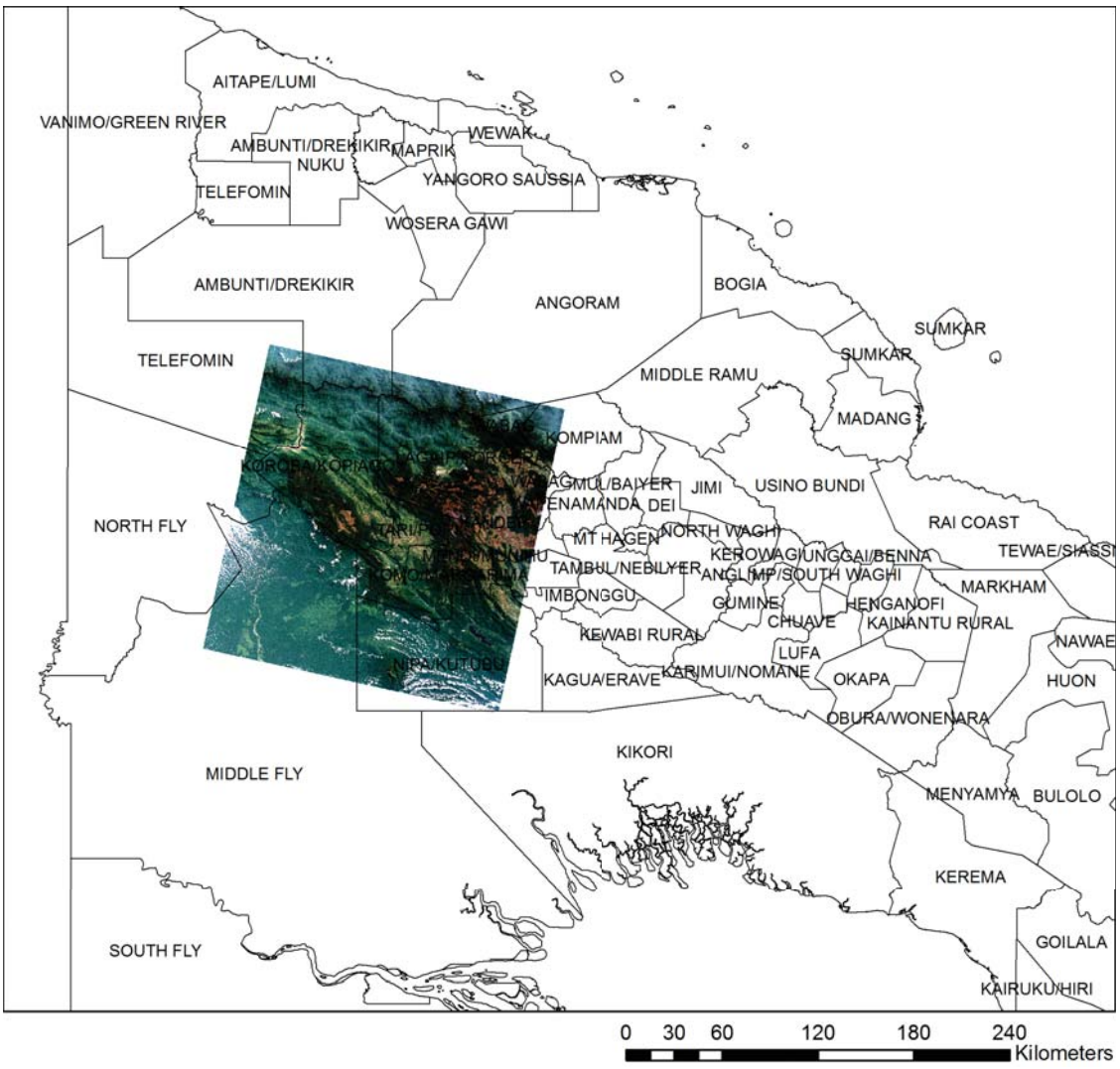

Figure 1. Study area positioned behind district boundaries.

3 Displaying the study area

Before discussing several displays of the study area, I will first shortly describe the key aspects of Landsat 8 imagery. Landsat 8 has been probing the earth since 2013, and thus provides a relatively new source of free image data. It has several upgrades with respect to the well-known Landsat 7 sattelite, including additional wavebands and a new Scan Line

Corrector(SLC) in the ETM+ instrument.[1] Table 1 in the Appendix shows an overview of the waveband properties of Landsat 8.

To better understand the study area, I present four displays: (i) a true colour composite, (ii) a pan-sharpened true color composite, (iii) a pan-sharpened standard false-colour near infrared composite and (iv) a pan-sharpened false colour composite specifically designed to help distinguish the road network from surrounding unpaved surfaces. The pan-sharpening transformation is based on a (weighted) average of the pan-chromatic band. The result of the (weighted) average is used to create a band adjustment value, which is used in calculating the adjusted output values for each band in the composition. The procedure is summarized using the following equation:

$$\hat{B}_n = B_n + A \tag{1}$$

Where the capitals represent $i \times i$ matrices with cross-sectional pixel data, $\hat{B}_n$ is the output for band n after applying the band adjustment A, to the original band data B_n which is calculated as:

$$A = P - \bar{P} \circ W \tag{2}$$

In which P is the pan-chromatic band, and $\bar{P}$ is it's respective average, possibly multiplied[2] by a weighting matrix W, which is constructed for an image with N channels as follows:

$$W = \frac{\sum_{n=1}^{N}(\beta_n B_n)}{\sum_{n=1}^{N}(\beta_n)} \tag{3}$$

Where β_n is the weight factor to band n, including the possibility for additional information included in an alpha band. Pan-sharpened bands are denoted throughout the text as (bands $\cdot$,$\cdot$,$\cdot$,8), not to be confused with a four-channel composition using an alpha band, which I denote (bands $\cdot$,$\cdot$,$\cdot$,$\cdot$,α).

The true color composite in figure 2 shows several roads intersecting the by forest dominated landscape. Some interesting features in the image include two air strips in the northeast and south, and a mining site near the center of the image. Immediately visible is that based only on the visible wavelengths, it is hard to distinguish artificial surfaces from the brownish surface that dominates the northeast corner of the map. The pan-sharpened image (ii) reveals a clear parcel-like structure, which might lead to think that these surfaces are temporarily barren croplands. Fortunately, Google Earth possesses high resolution

[1] The SLC in Landsat 7 failed, which caused the now infamous black stripes through images.
[2] Element by element multiplication, denoted using $\circ$ as the Hadamard product.

imagery of the area, which seems to support this idea.[3] The pan-sharpened image shows increased detail in small linear features, for example on the southern airstrip. The increased resolution of the pan-chromatic band will likely be key to succesful road recognition.

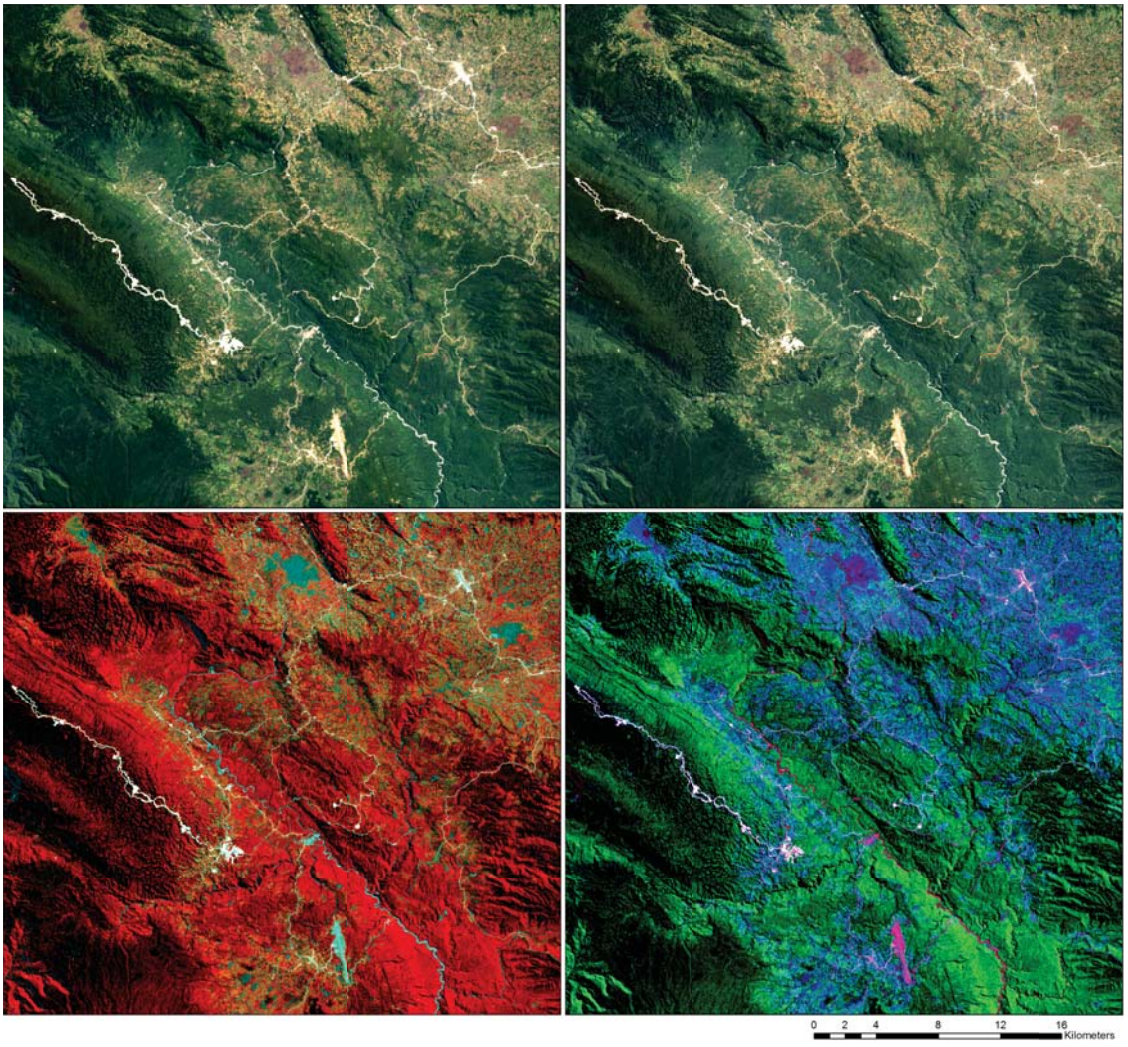

Figure 2. Displays of the central part of the image contained in Figure 1. (i) Top left) standard RGB composition (bands 4,3,2), (ii) Top right) pan-sharpened RGB (bands, 4,3,2,8), (iii) Bottom left) false color (bands 5,3,2,8), (iv) Bottom right) false color (bands 6,5,4,8).

Another interesting feature is the road that leads from the central mining site up into the mountain east in the image. This is a relatively new and frequently travelled industrial road. The image below is taken at ground level, and shows clearly that both the true

[3]Coordinates: $https://www.google.nl/maps/@-5.8462482,142.817213,794m/data=!3m1!1e3$

colour composition and the ground level photograph of the road show similar reflectance properties.[4]

Figure 3. Screen captured from Google Earth showing an agricultural landscape.

The basic RGB displays, and the ground photograph, reveal that unpaved road surfaces have a high reflectance throughout the visible wavelengths. Roads that are frequently travelled have higher reflectance than infrequently travelled surfaces. This could be potentially interesting for distinguishing roads from untravelled barren surfaces.

Figure 4. Spine Road to Wellpads, PNGLNG Project, Hides.

Further information on spectral differences between roads and visibly similar terrain, is revealed by the false colour compositions. The standard false colour composition (iii)

[4]Image source: *http* : *//www.panoramio.com/photo/*101618835

shows that rivers and roads differ in near infrared. Interesting is that the airstrips show up differently from road surfaces, while based on the RGB bands it is hard to distinguish them. Also the barren croplands in the north are more easily distinguished from road surfaces, though the difference is primarily due to differential in intensity of overall reflectance instead of specific wavelength properties. Finally, the alternative false colour composition (iv) reveals that roads differ substantially in shortwave infrared, which could be due to soil characteristics. The difference with rivers is notable and airstrips again seem to differ strongly from roads.

4 Spectral content analysis

To obtain more detailed insight in the spectral properties of road surfaces, I present radiance curves for two types of barren land and high altitude forest cover. All bands have been sampled at the specific locations presented in figure 5.

Mean DN values for each sampled area for all wavebands have been computed and are presented in figure 6. Mean DN values provide a clear indication on whether specific bands are useful in automated road mapping. However, to mathematically distinguish different surfaces, it is important that the mean DN values are efficient estimates for the DN values of the entire sampled area. Therefore I also analyze the Coefficients of Variation for DN values at different wavelengths.[5] These results are portrayed in figure 7.

All four types of land cover share a similar radiance pattern, though an important difference between forest cover and barren land types can be seen by comparing the black body temperatures that are consistent with figure 6. Radiance is average for the barren land types in the four visible wavelength bands and peaks in the near infrared band, consistent with a black body temperature of around 3300K.[6] Radiance then decreases again at shortwave infrared. The black body temperature of forest land cover, which peaks in the shortwave infrared frequencies, is much lower, around 1800K. Detail in spectral bands is thus insufficient for automated road recognition based on pixel-level black body temperatures, but it would be sufficient to seperate forest from barren surfaces.

[5] Also known as Relative Standard Deviations.

[6] This is seen by maximizing Planck's radiation Law w.r.t. the temperature T, at frequency $\lambda = \lambda_{max} 0.865\ \mu$m:

$$\arg\max_{T \in \mathcal{T}} 2\pi \mathrm{hc}^2/\lambda^5 \times 1/e^{\mathrm{hc}/\lambda \mathrm{kT}} - 1 \sim 3300$$

Where h is Planck's constant (6.266×10^{-34} Js); k is Boltzmann's constant (1.38×10^{-23} JK) c is the speed of light (3.10^8 ms^{-1}).

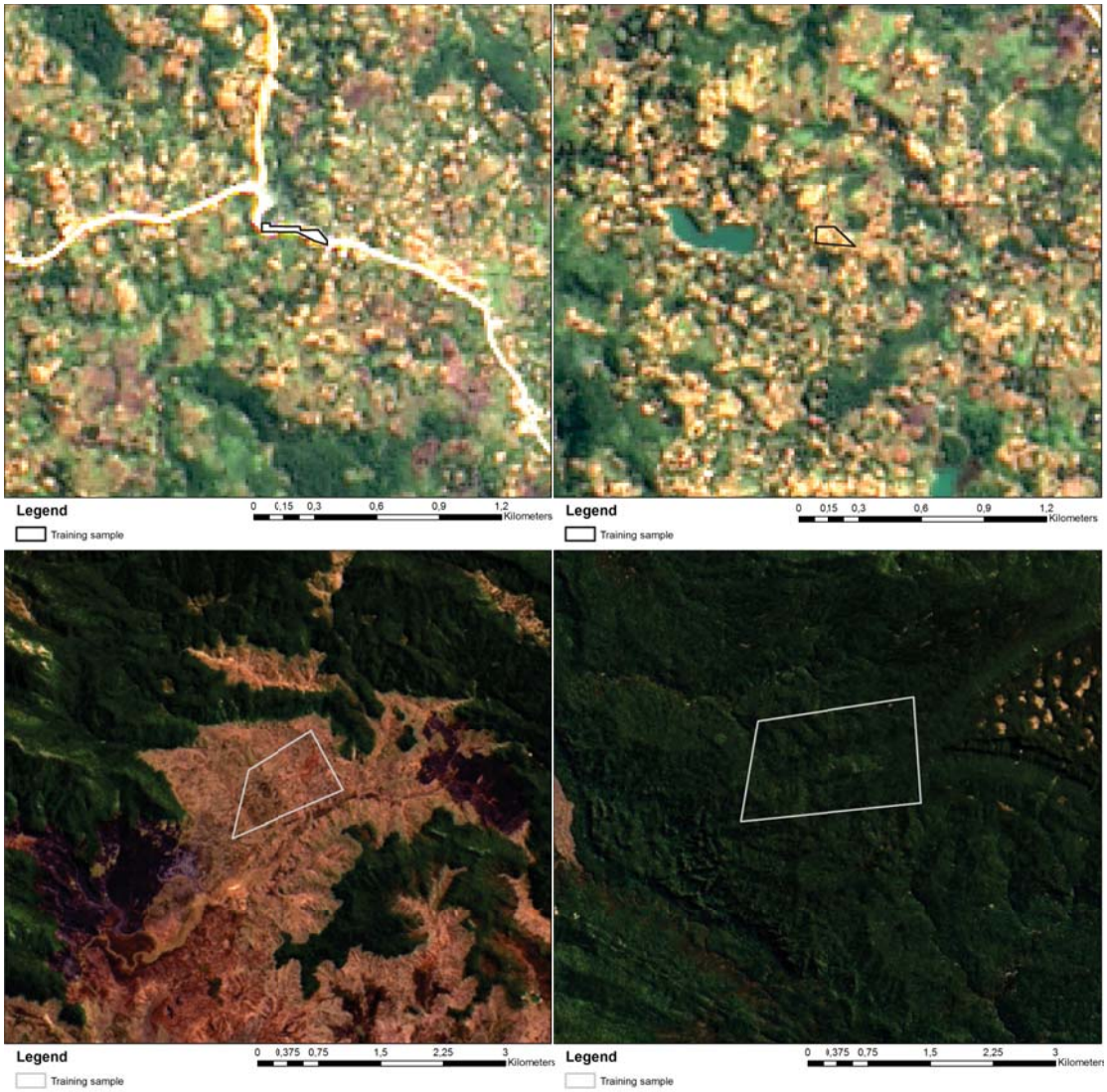

Figure 5. Sampled areas. (i) Top left) unpaved road. (ii) Top right) barren crop-land, indicated as barren type I. (iii) Bottom left) high altitude barren land cover, indicated as barren type II. (iv) Bottom right), high altitude forest cover. Pan sharpened RGB as background.

Several features from figure 6 are more promising for automated road detection. Road surfaces have higher radiance in all of bands 1-8. Furthermore, whereas for barren type I and II, and forest cover, radiance is higher in the blue band relative to the red band, the opposite is true for unpaved roads. Both findings are in line with the ground level pictures and true colour composites, in which road surfaces have a slight red/brownish tint. Mean DN in the pan-chromatic band summarizes this finding, and shows that roads show up brighter on average in the whole visible spectrum. As expected, all four classes have very

low and approximately equal mean DN in band 9, which is specifically designed to capture only cloud cover. This is due to the cloud- and mist free altitude at which the training sites have been sampled. Finally, all four land-cover types radiate relatively strong in the thermal infrared bands, which is to be expected in a tropical area. Roads and barren type I are quite similar in both thermal bands, whereas forests and high altitude barren type II radiate less longwave infrared, reflecting that these are colder areas. As thermal properties have a coarse spatial resolution, they are not likely very powerfull tools for road recognition.

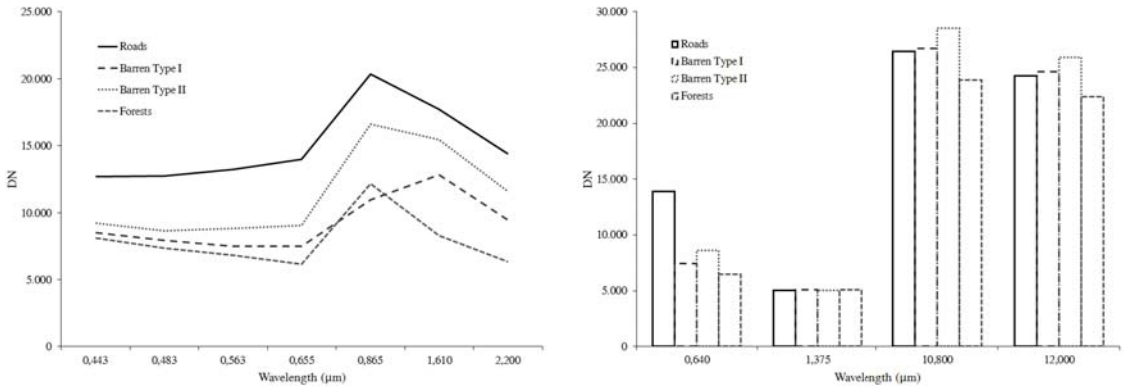

Figure 6. Left) Radiance curves for bands 1-7. Right) bars indicating radiance levels for bands 8-11. Figures bases on mean DN.

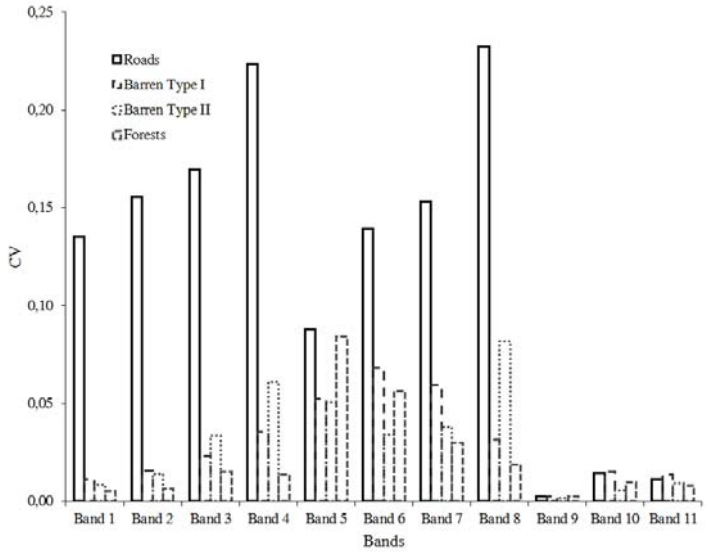

Figure 7. Coefficients of Variation for DN values for each band.

Based on mean DN values, it seems that roads are quite distinct from other land-cover types. Figure 7 however, conveys a slightly more nuanced message. While road surfaces strongly differ from other land-cover types in mean DN, these mean DN values are less representative for the entire road surface as a land-cover type. In all wavelengths, except for longwave infrared, variation in DN values is highest for roads.[7] Particularly important is that when only looking at mean DN, it seems that roads could be distinguished from other land-cover types based on radiance in the red band, while figure 7 shows that in this specific band variation is particularly high. This holds similarly for the pan-chromatic band. When looking at figure 6 and 7 combined, the most consistent band for distinguishing unpaved roads seems to be the first band, which is specifically designed to image shallow water and track fine particles contained in dust and smoke. The reason that these particles show up in the deep blue band is because blue is scattered by molecular bits of water or dust in the air. Similar as Rayleigh scattering causes the sky to look blue from ground level, from space surfaces like unpaved roads that send off lots of particulate matter show increased deep blue radiance. Finally, bands 9-11 have low variation in DN values, but from figure 6 we learned that mean DN values are close for all four land-cover types.

5 Ancillary data and high-pass filter analysis

In this section I investigate what we can learn from ancilarry data. Specifically, I focus on the positioning of roads in the geo-economic landscape to understand the typical locations where roads can be expected. I furthermore explore the potential of high-pass filters used on the high-resolution pan-chromatic band, to derive additional information helpful in road detection.

5.1 Roads in the geo-economic landscape

Figure 8 shows a larger overview of the study area, and includes population counts, road vector data and a hill shade made from ASTER elevation data. The area explored in the previous section shows up as the densely populated central part in figure 8. The figure reveals that there are several bush tracks, which are generally not seen on the satellite displays. Furthermore, ancillary data on roads confirms that visually distinguishable roads are indeed unpaved and that sealed roads are not apparent in this region. The hill-shade effect shows that the lower west part, which is also slightly misty, is flat, while the other part of the display constitutes mostly rugged terrain. This is important because low altitudes have more atmospheric distortion, and thus have different characteristics in the first wave

[7] The long wavelength infrared band has a 100m resolution, see table 1. This resolution is too coarse to capture variation in the thin road structures, accounting for the difference in CV.

band. While road detection in the Highlands region could be based on this particular band, detection in misty regions might require a different approach. Figure 8 also shows that roads closely follow the distribution of population counts, but this does not hold for the road in the southeast. Supplementing satellite data with detailed population data might not help road detection algorithms everywhere. However, the concentration of population and roads on high altitude terrain alleviates the slightly worrying notion that exploitation of the first wavebenad might have less power in the lowlands regions, as the main part of the network can be expected to intersect with higher altitude terrain.

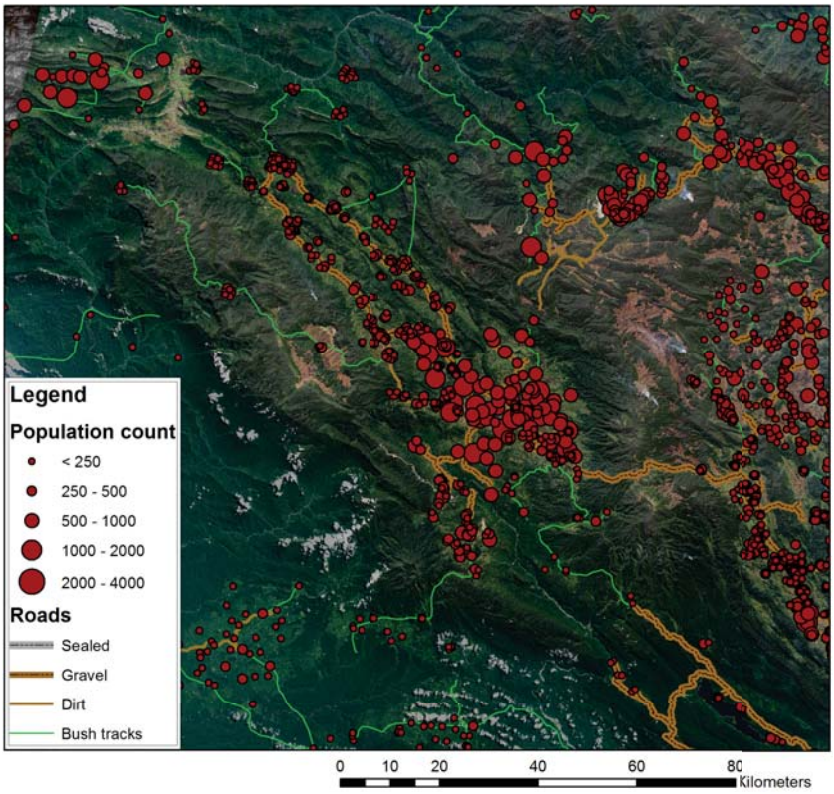

Figure 8. Map showing population density and road types. Pan-sharpened RGB display, enhanced with a hill shade, as background.

5.2 High-pass enhanced sharpening algorithm

Additional information to help distinguish roads can be obtained with high-pass filters. High-pass filters are known as edge-enhancing filters, and the particular filter used here is based on a 3×3 kernel to calculate the z-values w.r.t. neighbouring cells. For a single cell the procedure to obtain local z-values is:

$$z_{22} = \sum (A \circ K) \tag{4}$$

Where A is a 3×3 matrix of neigbhouring values:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{22} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{5}$$

And K is a 3×3 normalized kernel matrix:

$$K = \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{22} \\ k_{31} & k_{32} & k_{33} \end{bmatrix} = \begin{bmatrix} -\sqrt{d_{11}} & -\sqrt{d_{12}} & -\sqrt{d_{13}} \\ -\sqrt{d_{21}} & \sum_{i=1}^{3} \sum_{j=1}^{3} (\sqrt{d_{ij}}) & -\sqrt{d_{23}} \\ -\sqrt{d_{31}} & -\sqrt{d_{32}} & -\sqrt{d_{33}} \end{bmatrix} \tag{6}$$

Where d_{ij} is the Eucledean distance of each cell w.r.t. the centre:

$$d_{ij} = \|k_{ij} - k_{22}\| \tag{7}$$

Note that K always sums up to 0. Therefore, when A is a uniform matrix, the z-value computed for the centring cell is 0. When the outer elements of A are higher than the centre value, the z-value is negative, and for smaller outer values the result is positive. The z-value increases in magnitude when the difference between the centre value and outer values increases, such as is the case at sharp transitions in DN values.

To explore the potential of high-pass filters in automated road detection, I explore as a first step their potential to further enhance visualization of the road network. Figure 9 is generated by first computing z-values for the 15 m resolution pan-chromatic band using the procedure outlined above, and transforming z-values above a threshold of 2.5 standard deviations to 1, and 0 below that threshold.

$$\mathbf{z}_{22} = \sum (A \circ K)\, 1[(\sum (A \circ K)) > 2.5 \times \sigma(Z)] \tag{8}$$

Where $\sigma(Z)$ is the standard deviation of the $i \times i$ cross-sectional z-value matrix, and $1[Q]$ is the indicator function that takes a criterion Q.[8] The result is a binary cross-sectional matrix $\mathbf{Z}$ with elements $\mathbf{z}_{ij}$ that are nonzero for cells that have significantly higher DN value in the visual spectrum than their surrounding cells. This binary matrix is subsequently used as a fourth band in the weights equation (3), $B_4 = \mathbf{Z}$. The weights vector $\beta^{\intercal}$ for the four channel image is set to $(0.166, 0.167, 0.167, 0.5)$.

[8] The indicator function $1[Q] :\rightarrow [0, 1]$ is defined as $1[Q] = \begin{cases} 1 \text{ if } Q \text{ is } true \\ 0 \text{ if } Q \text{ is } false \end{cases}$.

Figure 9. Left) Standard RGB display. Right) an air strip in a high-pass enhanced image.

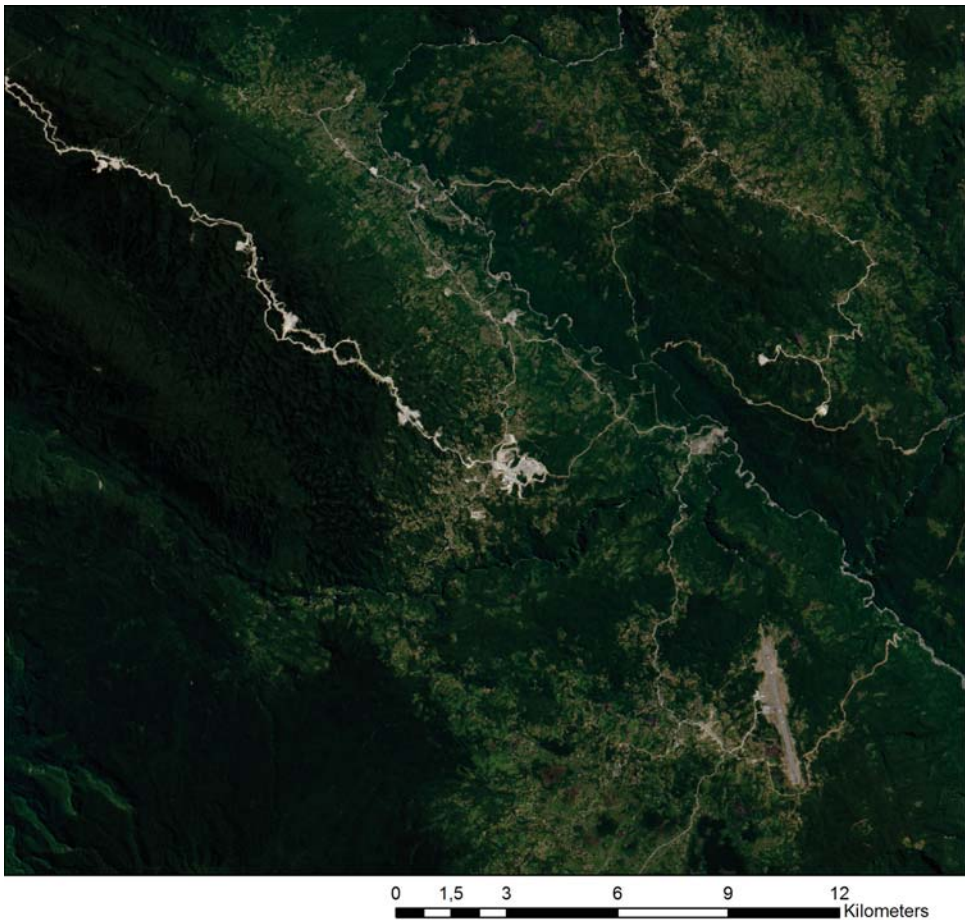

Figure 10. Overview of the high-pass filter enhanced image.

Figure 9 shows that this procedure is able to better reveal levels of detail by stabilizing the brightness and further accentuating the pan-chromatic sharpening around edges. Using the high-pass filter as a visual enhancer better reveals the detailed differences in pixel colour of the air-strip.

Figure 10 shows an overview of the high-pass filter enhanced image. The image is on average slightly darker due to the fact that DN values get less weight if they are not identified as edges by the high-pass filter. The image shows that this technique is able to show differential intensity of normally visibly similar pixels on unpaved surfaces such as the central mining area and the road that leads uphill.[9] For example, using figure 10 we can now distinguish the road uphill from washouts and see that there has been an old road, severely damaged by rainfall, and a newly constructed road.

6 Conclusion

In this short paper I explored wavelength properties of different land-cover classes to better understand how Landsat 8 imagery can be used for automated road recognition in developing countries with tropical conditions. Distinguishing roads from other barren surfaces constitutes the main challenge, and I showed that differences in spectral content are relatively small between barren surfaces while stronger w.r.t. forest cover. Specifically, the radiance curves have similar curvature for barren surfaces and as a result, derived black body temperatures have no power to distinguish different types of barren land. Detail in spectral bands of Landsat 8 is thus not sufficient to distinguish roads from barren surfaces. However, inspection of false colour composites revealed that in the Near Infrared and Short Infrared wavelengths, roads can be distinguished from other barren-like surfaces likely due to vegetation and soil characteristics. Generally, roads have stronger reflectance as measured by mean DN values throughout the whole visible wavelengths, but at the same time there is more variation in underlying DN values. The first waveband, designed to detect aerosols, has favorable characteristics for unpaved road detection. In this sepecific waveband, there is substantial difference in mean DN w.r.t. other barren-like surfaces while the CV is relatively low. However, tropical regions are plagued by atmospheric distortion and this useful property is likely only powerful for road detection at higher altitutes where atmospheric distortion is significantly less. In PNG this does not pose a great problem as both population and road occurrence intersects with high altitude terrain.

Based on these findings, I conclude that road recognition algorithms should rely on bands 1-8 and should initially focus on high altitude terrain to avoid atmospheric distortion.

[9] In both RGB and pan-sharpened RGB composites, there are no visible differences.

Reflectance in the deep blue spectrum should be a strong predictor of unpaved surfaces in general, due to high atmospheric concentration of dust particles. Feeding classification algorithms with elevation data, might solve problems related to increased reflectance in deep blue wave lengths in misty tropical valleys. Training samples for roads should be very carefully selected since there is large variation in spectral content. Training samples should not include air strips or washouts. Even though they are visibly similar in the RGB spectrum, these landscape features have different properties in the near and shortwave infrared wavelengths. Exploiting differences in these wavelengths is likely the key to further distinguishing between unpaved roads and other unpaved surfaces. Finally, I explored the potential of high-pass filters to enhance visualization of the road network. A boolean map derived from standard deviations of the z-values obtained with the high-pass filter, was used in the weighting equation of a four channel image. This stabilized brightness and accentuated the pan-sharpening around edges, revealing quite astonishing levels of detail. The output of the high-pass filter algorithm is likely to contribute further to predictive power of road classification algorithms.

References

Brecher, A., Noronha, V., Herold, M. (2004) UAV2003: A roadmap for deploying unmanned aerial vehicle (UAVs) in transportation, Findings of Specialist Workshop in Santa Barbara, CA, December 2003. Downloaded from: http://www.ncgia.ucsb.edu/ncrst/meetings/20031202SBA-UAV2003/Findings/UAV2003-Findings-Final.pdf (accessed: September 2015).

Chambon, S., Subirats, P., Dumoulin, J. (2009) Introduction of a Wavelet Transform Based on 2D Matched Filter in a Markov Random Field for Fine Structure Extraction: Application on Road Crack Detection. IS&T/SPIE Electronic Imaging, Image Processing: Machine Vision Applications II, San Jose, USA.

Cheng, H.D., Chen, J.R. , Glazier, C.,Hu, Y. G. (1991) Novel Approach to Pavement Cracking Detection Based on Fuzzy Set Theory. Journal of Computing in Civil Engineering 13 (4): 270-280.

Cheng, H.D., Wang, J.L., Hu, Y.G., Glazier, C., Shi, H.J., Chen, X.W. (2001) Novel Approach to Pavement Cracking Detection Based on Neural Networks, Transp. Res. Record 1764: 119-127.

Gecen, R., Sarp, G. (2008) Road Detection From High and Low Resolution Satellite Images. Commission PS-13, WgS IV/3.

Herold, M., Roberts, A. (2005) Spectral characteristics of asphalt road aging and deterioration: implications for remote-sensing applications. Applied Optics 44 (20):

4327-4334.

Mei, A., Salvatori, R., Fiore, N., Allegrini, A., D'Andrea, A. (2014) Integration of Field and Laboratory Spectral Data with Multi-Resolution Remote Sensed Imagery for Asphalt Surface Differentiation. Remote Sensing 2014 (6): 2765-2781.

Mohammadi, M. (2012) Road Classification and Condition Determination using Hyperspectral Imagery. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XXXIX-B7, 2012 XXII ISPRS Congress, 25 August – 01 September 2012, Melbourne, Australia.

Moussah, G., Hussain, K. (2011). A New Technique for Automatic Detection and Parameters Estimation of Pavement Crack. 4th International Multi-Conference on Engineering Technology Innovation, IMETI.

Zhang, C. (2008a) Development of a UAV-based Remote Sensing System for Unpaved Road Condition Assessment. Proc. of ASPRS Annual Conference. April 28 – May 2, Portland, OR.

Zhang, C. (2008b) Development of a UAV-based Remote Sensing System for Unpaved Road Condition Assessment. International Achieve of Photogrammetry and Remote Sensing. July 3-13, Beijing, China.

Zhang, C. (2010) Assessment of Rural Road Condition using UAV-based Remote Sensing. Proc. of Remote Sensing Technologies for Transportation Applications. Washington DC.

Zhang, C., Elaksher, A. (2012) Development of an Unmanned Aerial Vehicle-Based Imaging System for 3D Measurement of Rural Road Surface Distresses. Computer-Aided Civil and Infrastructure Engineering. Computer-Aided Civil and Infrastructure Engineering 27 (2): 118–129.

Appendix

Table 1: Overview of Landsat 8 wavebands

Spectral Band	Wavelength	Resolution
Band 1 - Coastal / Aerosol*	0.433 - 0.453 μm	30m
Band 2 - Visible Blue	0.450 - 0.515 μm	30m
Band 3 - Visible Green	0.525 - 0.600 μm	30m
Band 4 - Visible Red	0.630 - 0.680 μm	30m
Band 5 - Near Infrared	0.845 - 0.885 μm	30m
Band 6 - Short Wavelength Infrared	1.560 - 1.660 μm	30m
Band 7 - Short Wavelength Infrared II	2.100 - 2.300 μm	30m
Band 8 - Panchromatic	0.500 - 0.680 μm	15m
Band 9 - Cirrus*	1.360 - 1.390 μm	30m
Band 10 - Long Wavelength Infrared**	10.30 - 11.30 μm	100m
Band 11 - Long Wavelength Infrared**	11.50 - 12.50 μm	100m

* New with respect to Landsat 7. ** Included in Landsat 7 as a joint band at 60m resolution. Furthermore, all waveband frequency ranges have been adjusted to smaller intervals. DN precision has been upgraded from an 8-bit format to 16-bit format, thus now ranging from 0 through 65,535.